Psi Chi
Journal of Undergraduate Research

Interim Editor
Martha Zlokovich, PhD
Psi Chi Journal of Undergraduate Research
Psi Chi Central Office
P.O. Box 709
Chattanooga, Tennessee 37401-0709
Telephone: +1-573-651-9139
Fax: +1-423-265-1529
Email: martha.zlokovich@psichi.org

Publication and Subscription Office
Susan Iles
Managing Editor
Psi Chi Central Office
P.O. Box 709
Chattanooga, Tennessee 37401-0709
Telephone: +1-423-756-2044
Fax: +1-423-265-1529
Email: susan.iles@psichi.org
German Version of the Engagement With Beauty Scale

Isabella F. Dachs and Rhett Diessner*, Lewis-Clark State College

Recreational Ballet Students: The Mirror Image of Professional Ballet Aesthetics?

Jaime F. Kaplan, Towson University

Questionnaire Versus Performance-Based Functional Assessment in Dementia

Christine Karver, University of New Mexico; Terri Teshiba, Mind Research Network; Patrick Cordova, New Mexico VA Health Care System Research Service; John Adair**, Kathleen Haaland**, Joseph Sadek*, University of New Mexico

The Effect of Instructional Method on Second Language Acquisition: An Examination of Some Contributing Factors

Laura M. Morett, Washington College

Improving Bathroom Safety Among Community Dwelling Older Adults

Julia L. Murphy and Ann M. Steffen*, University of Missouri–St. Louis

Changes in Parents’ Stress as Their Children Become Adolescents: A Validation of the Stress Index for Parents of Adolescents

Melissa K. Wheatley and Diane E. Wille*, Indiana University Southeast

* Faculty mentors, **Faculty
About Psi Chi
Psi Chi is the International Honor Society in Psychology, founded in 1929 for the purposes of encouraging, stimulating, and maintaining excellence in scholarship, and advancing the science of psychology. Membership is open to graduate and undergraduate men and women who are making the study of psychology one of their major interests and who meet the minimum qualifications. Psi Chi is a member of the Association of College Honor Societies (ACHS) and is an affiliate of the American Psychological Association (APA) and the American Psychological Society (APS).

Psi Chi functions as a federation of chapters located at over 1,000 senior colleges and universities in the U.S. and Canada. The Psi Chi Central Office is located in Chattanooga, Tennessee. A Board of Directors, composed of psychologists who are Psi Chi members and who are elected by the chapters, guides the affairs of the organization and sets policy with the approval of the chapters.

 Psi Chi serves two major goals—one immediate and visibly rewarding to the individual member, the other slower and more difficult to accomplish, but offering greater rewards in the long run. The first of these is the Society’s obligation to provide academic recognition to its inductees by the mere fact of membership. The second goal is the obligation of each of the Society’s local chapters to nurture the spark of that accomplishment by offering a climate congenial to its creative development. For example, the chapters make active attempts to nourish and stimulate professional growth through programs designed to augment and enhance the regular curriculum and to provide practical experience and fellowship through affiliation with the chapter. In addition, the organization provides programs to help achieve these goals including regional and Society conventions, research award and grant competitions, certificate recognition programs, chapter awards, and Society service projects.

Journal Purpose Statement
The twofold purpose of the Psi Chi Journal of Undergraduate Research is to foster and reward the scholarly efforts of undergraduate psychology students as well as to provide them with a valuable learning experience. The articles published in this journal represent primarily the work of the undergraduate student(s). Faculty mentors, who deserve recognition, are identified by an asterisk next to their name or on a separate byline.

Because the articles in this journal are primarily the work of undergraduate students, the reader should bear in mind that: (1) the studies are possibly less complex in design, scope, or sampling than professional publications and (2) the studies are not limited to significant findings. The basis for accepting papers for publication is the agreement among three professional reviewers that the project, hypothesis, and design are well researched and conceived for someone with an undergraduate level of competence and experience.

Instructions for Contributors
The Psi Chi Journal of Undergraduate Research encourages undergraduate students to submit manuscripts for consideration. Submissions are accepted for review on an ongoing basis. Although manuscripts are limited to empirical research, they may cover any topical area in the psychological sciences.

1. The primary author of a submitted manuscript must be an undergraduate student who is a member of Psi Chi. Manuscripts from graduate students will be accepted only if the work was completed as an undergraduate student and not more than 6 months has passed since graduation. Additional authors other than the primary author may include non-Psi Chi students as well as the faculty mentor or supervisor. Membership verification information (member ID number) for the primary author must be included.

2. Only original manuscripts (not published or accepted for publication elsewhere) will be accepted.

3. All manuscripts must be prepared according to the Publication Manual of the American Psychological Association (5th ed.).

4. What to submit:
   a. A Microsoft Word electronic copy of the complete manuscript with figures, tables, and charts generated in either Word or Excel. Any scanned images or illustrations must be at least 600 dpi resolution. Make sure that identifying names, affiliations, etc. appear only on the title page and nowhere else on the manuscript (i.e., manuscripts should be reasonably free of clues to the identity of the authors). Footnotes that identify the author(s) should appear on a separate page.
   b. An email address so that receipt of your manuscript can be acknowledged.
   c. A sponsoring statement from the faculty mentor who attests:
      (1) that the research adhered to APA ethical standards;
      (2) that the mentor has read and critiqued the manuscript on content, method, APA style, grammar, and overall presentation; and
      (3) that the planning, execution, and writing of the manuscript represent primarily the work of the undergraduate student.

Submit all electronic files to: psijournal@psichi.org

The Psi Chi Journal of Undergraduate Research (ISSN 1089-4156) is published quarterly in one volume per year by Psi Chi, Inc., The International Honor Society in Psychology, P.O. Box 709, Chattanooga, TN 37401-0709.

Subscriptions are available on a calendar-year basis only (Spring–Winter). U.S. rates are as follows (four issues): Individual $20; Institution $40. For international rates or other information contact: Psi Chi Central Office, P.O. Box 709, Chattanooga, TN 37401-0709; telephone +1-423-756-2044; fax +1-423-265-1529; e-mail journal@psichi.org. Printed in the USA. Periodicals postage paid at Chattanooga, TN, and additional mailing offices.

Statements of fact or opinion are the responsibility of the authors alone and do not imply an opinion on the part of the officers or members of Psi Chi. Psi Chi does not accept paid advertising for its publications. Psi Chi Journal of Undergraduate Research. The society does not wish to appear to endorse any particular products or services. Copyright 2009 by Psi Chi, The International Honor Society in Psychology. Postmaster: Send address changes to Psi Chi Journal of Undergraduate Research, P.O. Box 709, Chattanooga, TN 37401-0709.
Copyright 2009 by Psi Chi, The International Honor Society in Psychology (Vol. 14, No. 3/ISSN 1089-4136)

In beauty (happily) I walk.
With beauty before me I walk.
With beauty behind me I walk.
With beauty below me I walk.
With beauty above me I walk.
With beauty all around me I walk.
It is finished (again) in beauty.

(Navajo/Diné Prayer; quoted in Underhill, 1956, p. 212)

Since ancient times, people have been fascinated by the beautiful. The earliest theories of beauty, in the West, were developed by the ancient Greeks. One such theorist was Pythagoras, who placed great emphasis on the golden ratio, which in mathematics, refers to a geometric proportion in which a line is divided. This ratio appears in objects humans experience as more beautiful, thus intertwining both mathematics and beauty. In fact, “since antiquity many philosophers, artists, and mathematicians have been intrigued by the golden section, which Renaissance writers called the divine proportion. It is widely accepted that a rectangle with sides in this ratio exhibits a special beauty” (Golden section, 2009). Others, such as Plato (1937; viz. The Republic, Book VI, 505b-508e) focused on the concepts of beauty, truth, and good —especially the unity among these three ideas.

Beauty has always been a topic of interest in philosophy, especially in the philosophical discipline of aesthetics. Consider, for example, Kant (1790/1987) who concentrated on both objective and subjective qualities of beauty. Among his most famous statements regarding beauty are: “The judgment of taste, therefore, is not a cognitive judgment, and so not logical, but is aesthetic which means that it is one whose determining ground cannot be other than subjective” (p. 44) and, when referring to the observer, “he judges not merely for himself, but for all men [sic], and then speaks of beauty as if it were a property of things” (pp. 55-56).

Even though there is a paucity of research with regard to modern psychological theories of beauty, with most of it focusing on visual arts (Eysenck, 1988; Oatley, 2002) or music (Gabrielsson & Juslin, 2003), research from humanistic psychology needs to be

German Version of the Engagement With Beauty Scale

The Engagement with Beauty Scale (EBS) was designed using the philosophical works of Kant (1790/1987), Hegel (c. 1835/1993), and Aquinas (c. 1260/1947), as well as the psychological work of Haidt and Kilter (2004). The scale has construct, concurrent, and predictive validity as well as internal consistency, with a Cronbach’s alpha of .90 for total score with an American sample (Diessner, Solom, Frost, Parsons, & Davidson, 2008). In the study reported here, the EBS was translated into German and administered to a sample of Germans living in southern Germany, revealing a Cronbach’s alpha of .94 for total score (validity was not formally assessed). Discussion centers on comparing the German sample’s scores on the natural beauty, artistic beauty, and moral beauty subscales of the EBS with the American sample’s scores.
acknowledged. For example, an important component of Rogers’ (1961) theory of the fully functioning person involves openness to experience which he defined as the opposite of defensiveness. In his model of how to acquire a rich and engaged life, Rogers also gave special attention to beauty and the emotional reactions that can be elicited from it when we are fully immersed in the moment, interacting and engaging with the environment.

Another prominent representative of the field of humanistic psychology was Maslow, most known for his hierarchy of needs (1943). He (1964) wrote extensively about awe or peak experiences, distinguishing between peakers and nonpeakers, and he also focused on human virtues. Maslow (1970) considered beauty to be a B(eing)-value (meta-motivation); that is, part of the B(eing)-cognition which motivates us for psychological growth and the development of our full potentials.

With regard to psychosocial correlates of engagement with beauty (i.e., cognitively appreciating and at the same time being emotionally involved with beauty), there is still a considerable void in scientific literature. However, there is a significant positive correlation between spiritual transcendence (Piedmont, 1999) and appreciation of beauty. This relationship had been speculated upon by Haidt and Keltner (2004) before having been empirically demonstrated by Diessner et al. (2008). Furthermore, appreciation and gratitude also are correlated (McCullough, Emmons, & Tsang, 2002).

It was not until the positive psychology movement’s (Peterson, 2006; Peterson & Seligman, 2004; Snyder & Lopez, 2002) recognition of the importance of appreciation of beauty as a character trait that the Engagement with Beauty Scale (EBS) was developed (Diessner; Rust, Solom, Frost, & Parsons, 2006; Diessner, et al., 2008). This scale constitutes a direct response to Haidt and Keltner (2004), who noted the lack of an instrument measuring this particular construct. The EBS is a 14-item self-report scale describing various levels of cognitive and emotional engagement concerning natural, artistic, and moral beauty. Theologically, “moral beauty” can be seen wherever the attributes of God are manifest, or, philosophically and psychologically, when we witness acts that involve human virtues. According to Peterson and Seligman (2004), these virtues can be grouped into the following six universal clusters: wisdom and knowledge, courage, humanity (love and kindness), justice, temperance, and transcendence—appreciation of beauty and excellence belonging to this last cluster. This “engagement” then may result in “self-transcendent emotions such as awe, admiration, and elevation” and various bodily responses (e.g., “the proverbial lump in the throat,” (Haidt & Keltner, 2004, p. 539).

The scale is based on the works of Kant (1790/1987; nature), Hegel (c. 1835/1993; art) and Aquinas (c. 1260/1947; morality and virtues) as well as on the psychological work of Haidt on moral beauty and the moral emotion of elevation (Haidt, 2000, 2002, 2003, 2006; Haidt & Keltner, 2004; Keltner & Haidt, 2003). Although there are two other instruments which measure appreciation of beauty as a subscale—the 10-item Appreciation of Beauty and Excellence (ABE) subscale of the 240-item Values in Action Inventory of Strengths (VIA-IS; Peterson & Seligman, 2004) and the Openness to Aesthetics (McCrae, 1996) subscale of the revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992)—the EBS is the only existing full scale that is devoted to measuring the trait of engagement with beauty.

In particular, the EBS subscale of moral beauty was modeled after Haidt’s (2002) work on elevation (the moral emotion that is brought about by appreciation and appraisal of beauty) and prosocial behavior associated with this emotion. According to Haidt (2002), moral emotions are “emotions that respond to moral violations or that motivate moral behavior” (p. 853). He (2002) further states that elevation “is elicited by moral beauty [and] appears to be caused by seeing manifestations of humanity’s higher or better nature; it triggers a distinctive feeling in the chest of warmth and expansion; it causes a desire to become a better person oneself; and it seems to open one’s heart, not only to the person who triggered the feeling, but also to other people” (p. 864).

Given this theoretical background and the relative paucity of cross-cultural investigations into the topic of engagement with beauty, the aim of the present research was to develop an internally consistent and reliable German translation of the EBS, a first step in establishing cross-cultural validity of the scale. Without a German translation, no reliable data can be collected from native speakers of German. It should be noted that German is not only the first language in Germany, but also in Austria, the Principality of Liechtenstein, and parts of Switzerland, Italy, and Belgium. Therefore, German is the most widespread first language in Europe, spoken by about 100 million people (Auswaertiges Amt, 2006).

The examination of engagement with beauty in different cultural contexts is timely because the positive psychology movement has clearly established the importance of this particular character trait and its relevance for counseling and positive psychotherapy. More empirical evidence concerning levels of engagement with natural, artistic, and moral beauty among people from different backgrounds would enhance our understanding of the interplay between culture
and this specific character trait.

Furthermore, because the previous American study relied on a convenience sample comprised exclusively of white undergraduate students at a small college in the Northwest, a further objective was to improve the generalizability of findings by testing a more diverse sample in the present study.

Method

Participants
A convenience sample of N = 71 Germans completed the measure, two of whose protocols were unusable. The N = 69 consisted of 59.4% women, ages 18 to 81 years (M = 48.8; SD = 16.6). The sample was 98.6% Christian and 1.4% Jewish; nationality was 100% German. At time of data collection, all participants lived in or near the town of Memmingen, Bavaria, in southern Germany.

Materials
The EBS uses a 7-point Likert scale ranging from “very unlike me” to “very much like me” on items such as “When perceiving beauty in nature I feel changes in my body, such as a lump in my throat, an expansion in my chest, faster heartbeat, or other bodily responses”; “When perceiving beauty in a work of art I feel something like a spiritual experience, perhaps a sense of oneness or being united with the universe or a love of the entire world”; and “When perceiving an act of moral beauty I find that I desire to become a better person” (Diessner et al., 2008).

The EBS provides a total scale score (possible range 14–98) and also has three subscales tapping engagement with natural beauty (four items; possible range 4–28), artistic beauty (four items; possible range 4–28), and moral beauty (six items; possible range 6–42). A study of the EBS with an American sample (N = 206; 58% female) yielded a total score internal consistency of = .90 and subscale alphas of .80 for Natural Beauty, .87 for Artistic Beauty, and .85 for Moral Beauty (Diessner et al., 2008).

In an American study (N = 206), the EBS showed adequate internal consistency and temporal stability, with a total score of .90–.91, and test-retest reliability from .79–.85; the EBS Natural Beauty subscale was .80; Artistic Beauty subscale from .87–.88; and the EBS Moral Beauty subscale from .85–.89. After confirmatory factor analysis, a chi-square test and a root-mean-square error of approximation (RMSEA) demonstrated a satisfactory fit for the model with three correlated subscales of the EBS; likewise, several goodness of fit indices (AGFI, GFI, CFI), ranging from .91–.97, indicated the 3-factor model of the EBS as reasonable (Diessner et al., 2008).

The EBS has a positive correlation of .80 with the Appreciation of Beauty and Excellence subscale of the Values in Action Inventory of Strengths (the ABE of the VIA-IS; Peterson & Seligman, 2004; note that this subscale of VIA-IS measures appreciation of both beauty and excellence, whereas the EBS is solely devoted to engagement with beauty). As predicted by Haidt and Keltner (2004), concurrent validity studies show the EBS has medium to high positive correlations with measures of gratitude and spiritual transcendence and low negative correlation with materialistic values. A known-groups study demonstrated that the EBS Artistic Beauty subscale differentiated students engaged in the arts from those who are not (Diessner et al., 2008).

In the present study, the American-English EBS was translated into German by the first author who was helped by an independent translator (a German college-preparatory teacher of English and French and lecturer in adult education specializing in business English, possessing a German university degree equivalent to a M.Ed.) who otherwise did not participate in the study. The scale was then back-translated by another German college-preparatory teacher of English and French, also possessing a German university degree equivalent to a M.Ed. (who was not further involved in the study either), and compared to the original. The back-translation was judged by both translators to compare favorably to the original English one. Both translators agreed on its validity and therefore there was no need to alter the initial translation into German.

Procedure
The Institutional Review Board at the first author’s US college approved the study, informed consent was obtained from all participants, and debriefing was offered. Using her social network in her hometown of Memmingen, Bavaria, southern Germany, the first author of this study contacted 71 individuals, approached them in person, and asked them to complete the translated self-report measure during the summer of 2007.

Specifically, prior to distributing this sole measure, participants were verbally informed of the purpose of
the study, and it was emphasized that they were free to participate or not and that they could stop at any time if a questionnaire statement unduly upset them. In order to avoid procedural bias or any influence on participants’ attitudes and task performance, which might lead to socially desirable responses and a distortion of results, no monetary or other type of incentives were provided. Participants were further instructed that their anonymity would be guaranteed, that the completed measures would be stored in a safe place, and that only the principal investigator and her research supervisor would have access to the completed measures.

No time limit was given for returning the questionnaire, and participants were free to complete it on site (usually in the comfort of their own homes) or to return it by post in a prepaid envelope to the first author’s home address in Germany. Of the 91 questionnaires distributed, 71 were returned to the first author (response rate of 78%). However, due to incompleteness, 2 questionnaires had to be excluded, resulting in 69 usable protocols.

Results

The German participants’ responses yielded Cronbach’s alphas of .94 for EBS total score; .85 for the Natural Beauty subscale; .85 for the Artistic Beauty subscale; and .90 for the Moral Beauty subscale. The EBS total beauty score can range from 14 to 98, and in this sample ranged from 21 to 92 (M = 61.5, SD = 17.7). Both the Artistic Beauty and the Natural Beauty subscales have possible ranges of 4–28 with ranges for both in the present study of 4–27. Artistic Beauty subscale was M = 17.2 (SD = 5.9); Natural Beauty subscale was M = 18.8 (SD = 5.6). The possible range for the Moral Beauty subscale is 6–42 with a range of 6–40 in the present German sample (M = 25.4, SD = 8.4).

As Table 1 shows, comparing the German EBS scores to the original American sample’s scores revealed that the American average was higher, t(273) = 3.26, p < .001, two-tailed; with an effect size of d = .45. There was no significant difference between Germans and Americans on the Artistic Beauty subscale. On the Natural Beauty subscale, the Americans scored higher, t(273) = 2.61, p = .01, two-tailed, with effect size d at .36; and on the Moral Beauty subscale, the American average was also higher at t(273) = 5.6, p < 0.001, two tailed, with an effect size of d = .78.

Discussion

Our objectives were to develop a reliable German translation of the Engagement with Beauty Scale and to subsequently administer it to a sample of Germans. Based on the relatively high Cronbach’s alpha of .94 for total score, the German EBS appears internally consistent and reliable. The alphas on the Natural Beauty and Moral Beauty subscales of the German EBS were higher on those two subscales than was the original American sample, as was the German EBS total score alpha.

As can be seen in Table 2, it is difficult to explain and understand the Germans’ significantly lower scores on the EBS total score (M = 61.5, SD = 17.7) compared to the Americans (M = 68.7, SD = 15.5); as well as on the Natural Beauty subscale (German M=18.8, SD=5.6; American M = 20.7, SD = 5.1), and the Moral Beauty subscale (German M = 25.4, SD = 8.4; American M = 31.4, SD = 7.4). On the one hand, it is possible that these are real differences, and that Americans—at least the college students tested—actually do perceive themselves as more engaged with beauty than Germans do. On the other hand, Germans might have a higher standard for engagement with beauty, and thus when they mark a “5” on the EBS’s 7-point Likert scale, they...
may be indicating higher engagement with beauty than when Americans mark a “6” in this context.

An important limitation of the present study is the relatively small sample size. Furthermore, both the previous study using the original English EBS and the new study conducted in Germany with the German version of the scale solely relied on nonrandom convenience samples (i.e., both samples were not representative of their nations as a whole).

Therefore, it is not possible to make generalizations. Future studies will need to examine the temporal stability of the newly-developed translation of the measure. In particular, it will be important to employ multi-method assessments using more diverse subject samples to investigate the temporal course of how people engage with various forms of beauty.

Another caveat is that, just as with the original English EBS, this new German version constitutes a subjective self-report measure. Thus, it is unclear whether participants accurately responded to each item or only gave a socially desirable response, which ultimately raises concern about the validity of the findings. It might be advisable to include a “lie scale” in a revised version of the scale.

Last, the present study lacks validity data. Although the results imply that the translation makes good sense to native speakers of German, validity was not formally assessed because the focus was on prerequisites of cross-cultural validity. In a future study, however, it will be fruitful to make up for this and therefore to also focus on obtaining validity data.

Despite the limitations of the initial study, this measure may become a potentially useful tool in basic and clinical research. In addition to the present German version, the EBS has also been translated into Cypriot Greek, Farsi, Croatian, and data have been collected in Cyprus, Iran, Croatia, and American Samoa (Richel et al., 2008). The Cronbach’s alphas have all been in the high .80s to low .90s for EBS total score, so it appears that the conceptualization of engagement with beauty across natural, artistic, and moral dimensions provides a coherent construct in a wide variety of cultures.

Engagement with beauty may have important implications in counseling psychology, as shown by recent research in positive psychology indicating that appreciation of beauty may be one of the most important and effective character strengths involved in overcoming depression and other psychological disorders (Seligman, 2002; Peterson, Park, & Seligman, 2006). It seems that facilitating positive change in psychotherapy will be the key because, “[t]he symptoms of depression often involve lack of positive emotion, lack of engagement, and lack of felt meaning, but these are typically viewed as consequences or mere correlates of depression. We suggest that these may be causal of depression and therefore that building positive emotion, engagement, and meaning will alleviate depression.” (Seligman, Rashid, & Parks, 2006, p. 775)

References

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Beauty</td>
<td>American</td>
<td>206</td>
<td>20.66</td>
<td>5.07</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>69</td>
<td>18.77</td>
<td>5.57</td>
</tr>
<tr>
<td>Artistic Beauty</td>
<td>American</td>
<td>206</td>
<td>16.67</td>
<td>5.92</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>69</td>
<td>17.25</td>
<td>5.88</td>
</tr>
<tr>
<td>Moral Beauty</td>
<td>American</td>
<td>206</td>
<td>31.42</td>
<td>7.39</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>69</td>
<td>25.45</td>
<td>8.35</td>
</tr>
<tr>
<td>Total Beauty</td>
<td>American</td>
<td>206</td>
<td>68.75</td>
<td>15.48</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>69</td>
<td>61.46</td>
<td>17.73</td>
</tr>
</tbody>
</table>


Kant, I. (1853). *Critique of judgement* (W. Pluhar, Trans.). Indianapolis, IN: Hackett. (Original work published 1790)


A

eroxia. Common associations with this word are bones, starvation, and pathology. To ballet dancers, these familiar associations are often hidden by what they perceive as the more important concerns of success, beauty and perfection. The high prevalence of eating disorders seen in ballet dancers is not a new phenomenon; however, there is more to this correspondence than what lies on the surface. Ballerinas need more than grace and perfect technique in order to be successful in their career; they also need to be particularly thin. The thin-ideal pressures become stronger in the ballet world as the pressures in the larger society increase (McPhee, 2000).

Definitions of Eating Disorders
The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR), recognizes three types of eating disorders: anorexia nervosa, bulimia nervosa, and eating disorder, not otherwise specified (American Psychiatric Association, 2000). Within these major categories lie subcategories as well.

Anorexia nervosa. According to the DSM-IV-TR, anorexia nervosa is identified by an intense fear of gaining weight, refusal to maintain a minimally normal weight, a disturbance in one’s body perception, the absence of at least three consecutive menstrual cycles, and weight loss by a reduction in food intake or purging (APA, 2000). This weight loss or failure to gain weight leads to a continuance of body weight at less than 85% of what is considered typical for one’s height. By dividing weight in kilograms by height in meters squared, an individual can determine her body mass index (BMI). A BMI below 18.5 is considered underweight, although a BMI below 17.5 is a diagnostic criterion for anorexia nervosa (APA, 2000).

Bulimia nervosa. Although bulimia nervosa varies from anorexia nervosa, one main underlying force feeds both disorders—control. While people with anorexia nervosa compensate for their feelings of loss of control by exerting too much control over their food intake, those with bulimia nervosa actually lose control during a binge. A binge is defined as eating a larger than normal amount of food in a discrete period of time (APA, 2000). In order to erase the effects from the binge, one may engage in some type of compensatory behavior, with 80–90% of those with bulimia nervosa using self-induced vomiting. Binge eating and compensatory behaviors must occur on average two times a week for at least three months in order to meet diagnostic criteria for bulimia nervosa.

Eating disorder, not otherwise specified. The largest category of eating disorders—eating disorder, not otherwise specified (APA, 2000).

The present study investigated body dissatisfaction, drive for thinness, perfectionism, and asceticism in a sample of recreational ballet students and preprofessional ballet students. We questioned whether the differences between the two groups in weight pressures would be reflected in how students perceived themselves. Fifty-five ballet students (23 recreational and 32 preprofessional) completed the Body Image Silhouettes subscale of the Kids Eating Disorder Survey, the Eating Disorder Inventory-3, and a demographics questionnaire. Results indicate the recreational ballet students did not differ significantly from the preprofessional ballet students. Both groups of dancers reported an ideal ballet body image that was significantly smaller than their ideal body image. Recreational ballet students, however, reported a significantly larger perceived body image than preprofessional ballet students.

Recreational Ballet Students: The Mirror Image of Professional Ballet Aesthetics?

JAIME F. KAPLAN
Towson University

JAIME F. KAPLAN
Towson University

Author Note. The author appreciates the help from her advisor and mentor, Dr. Jean F. Ayers. She would also like to thank Christopher Magalis for his assistance with the statistical portion of this study.
otherwise specified (EDNOS)-includes all clinically significant disorders of eating that do not meet the complete diagnostic criteria for anorexia nervosa or bulimia nervosa, as indicated by the DSM-IV-TR (Keel, 2005). A form of EDNOS that pertains to athletes is called anorexia athletica. Athletes with anorexia athletica tend to have a better prognosis than eating disordered nonathletes. The findings are such because this type of eating disorder is more focused on the behavioral level rather than on a deeper level of psychopathology (Bachner-Melman, Zohar; Elstein, Elizur, & Constantini, 2006).

With a high emphasis on weight loss, it is clear how having an eating disorder impairs women’s ability to see their bodies clearly, as most of these women see themselves as fat or overweight when they actually are not. Body image can be defined as the internal representation of one’s own perception of one’s outer appearance. Feingold (as cited in Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999) conducted a meta-analysis of attractiveness research. He specifically evaluated the subjective versus the objective nature of attractiveness. Feingold concluded, with regard to women, that only 6% of their assessment of appearance was determined by actual attractiveness. The other 94% of how women saw themselves was consequently due to outside influences, such as the media, culture and other people’s opinions (Thompson et al., 1999).

**Eating Disorders In Ballet**

Thinness was not always of great importance in the ballet community (Thompson et al., 1999). George Balanchine, founder of the New York City Ballet, set the standards for the ideal ballet look over 60 years ago that are still in place today. These standards are based on a straight body-long limbs, long neck, flat chest, toned muscles, good proportions-what some may refer to as the anorexic look or otherwise known as the prepubertal look (Ackard, Henderson, & Wonderlich, 2004). Balanchine wanted to see the bones in his dancers, and this was often the way he explained his image to these women (Thompson et al., 1999). Women with a delay in menarche are more likely to achieve Balanchine’s perfect long and lean body, as opposed to women with early or on time maturation (Hamilton, Brooks-Gunn, & Warren, 1986). Although some girls are genetically predisposed to mature later than others, young girls can influence the time at which they reach menarche by reducing their body fat composition through disordered patterns of eating.

In the ballet world it is the other dancers, teachers, and role models who promote the ballet ideal on a daily basis. They surround young girls aspiring to become ballet dancers, and are especially influential. Due to the high respect and/or fear dancers have for their teachers, dancers are willing to do what they are told and comply with whatever regimen they are faced with in order to achieve the ideal body type (Benn & Walters, 2001). Thinness in the dance world is not equivalent to thinness in the larger society. Dancer thin is comparable to the diagnostic criterion of anorexia nervosa, with the percentage of body fat being almost identical. Although most people in the dance world can attest to the prevalence of disturbed eating, it still remains dancers’ “dirty little secret” (McPhee, 2000).

The pressure to achieve the dancer-thin ideal is not only placed on professional adult dancers. Weight concerns in aesthetic sports are prevalent in younger children (Bachner-Melman et al., 2006). In a study by Ackard, et al. (2004), children as young as age 5 years who were dancers scored higher on a subscale for weight concern than their nondancer peers. These children understood the behaviors and beliefs that surround them and internalized these messages as something that needed to be done to fit in. For example, a 12-year-old dancer in the documentary, Dying To Be Thin, revealed that she was told to lose weight and comprehended that when she was thinner, her teachers nurtured her, gave her more attention and awarded her with better roles. She further reported that when she started gaining some of the weight back, she felt ashamed, weak, and like a failure (McPhee, 2000). In fact, adolescent ballet dancers often view themselves as less desirable, less attractive, and less confident than their peers who are not dancers (Price & Pettijohn, 2006). Not surprisingly, adolescent ballet dancers are dramatically more at risk for developing an eating disorder than their peers who are not ballet dancers (Thomas, Keel, & Heatherton, 2005).

**The Present Study**

The dance community as a whole may be the most obsessively weight conscious subculture in the nation. The present study investigated body perception to determine if female adolescent ballet dancers shared similar perceptions of the ideal ballet body, as well as how this perception affected the students’ perception of themselves, both in the ballet culture and larger society. It further compared body image perceptions and weight concerns between adolescent dancers at two different levels of involvement in the art: recreational and preprofessional ballet students. The study examined whether the differences in weight pressures from dance teachers and directors across studio types would influence how students perceived themselves. Last, this study assessed characteristics prevalent in both people with eating pathology and ballet dancers (e.g., perfectionism and drive for thinness).
Method

Participants
Fifty-five ballet students from six dance studios in the metropolitan area of a large East Coast city completed this study. Dance studio owners were contacted via e-mail and telephone with a description of the study and were asked if they had an interest in participating. Once permission was granted, directors’ signatures were collected and parental consent forms were handed out to all students eligible for participation. Inclusion criteria for the study were: women, age range 11–17 years, and participation in at least one ballet class per week. Involvement in the study was voluntary and required parental consent because all participants were minors. The average age of the participants was 14.3 years. The majority of participants, 54.5%, were Caucasian, while 34.5% were African American, 7.3% Hispanic, and 3.6% “other.”

The comparative groups included 32 ballet students from preprofessional ballet studios and 23 dance students from recreational dance studios. For the purposes of this study, dance studios were preclassified as preprofessional or recreational based on criteria generally established in the literature on dance education (Thomas et al., 2005). Preprofessional studios were identified as having a minimum requirement of three or more ballet classes per week and placing a strong emphasis on technique and terminology. A dance uniform, such as a specific color leotard per technique level, is typically required for students of this type of studio. At the highest levels of preprofessional training, auditions are required. Recreational studios were defined as those that offer many forms of dance, with no required minimum number of ballet classes per week. These studios stress having fun as a priority over mastering technique.

Materials
Testing materials included an eating disorder inventory and a body image scale. Participants also answered a demographics survey, which included questions about dance history, perception regarding the ideal ballet body, and family and personal history of eating difficulties.

Eating Disorder Inventory-3. The EDI-3 (Garner, 2004) is a self-report standardized measure that aids in the assessment of various areas of eating pathology. Although there are 12 scales in total, this study focused on the four scales of Drive for Thinness, Body Dissatisfaction, Perfectionism, and Asceticism. There are 91 items on the measure, to which participants use a Likert scale of 0–4 for responding.

As a requirement of the EDI-3, participants’ height and weight (collected by self-report of the participants) were used for the computation of the participants’ Body Mass Index (BMI). BMI was calculated after the collection of materials, without the participant present.

Kids Eating Disorder Survey. The Body Image Silhouettes subscale of the Kids Eating Disorder Survey (KEDS; Childress, Brewerton, Hodges, & Jarrell, 1993) is used to distinguish perceptions of body image. The subscale presents 8 images of girls in bathing suits in varying shapes and sizes. Each image is labeled from 1 to 8, with higher numbers representing larger body sizes, and the lowest number representing the thinnest, almost skeletal-like image. Instructions asked participants to circle the body image with which they most identified themselves and to underline the body image that they believed to be their ideal body type. An extra requirement, specific to this study, requested participants to place a “B” over the body image that they felt represented the ideal ballet body.
Procedure
Administration of all three surveys occurred in one, 30-min session at the participants’ studios. Each participant received an envelope containing all testing materials. Each envelope had a number assigned to it, with all subsequent materials marked with the same number. Participants did not write their names anywhere on the testing materials to ensure the anonymity of their responses.

Results

Descriptive Statistics

Eating Disorder Inventory-3. Participant reports indicated that 13% of the recreational dancers and 12.5% of the preprofessional dancers endorsed that they may have an eating disorder and/or are worried about their eating habits. Frequency data were analyzed for four subscales of the EDI-3: Drive for Thinness, Body Dissatisfaction, Perfectionism and Asceticism. Categories for scores ranged from common to rare, with rare scores indicating a disturbance in that specific area (See Table 1). Analysis was specific to these four subscales based on their direct relevance to characteristics of many ballet dancers (McPhee, 2000). Possible score ranges for the four subscales are as follows: Drive for Thinness: 0–28, Body Dissatisfaction: 0–40, Perfectionism: 0–24, and Asceticism: 0–28. Higher scores on these subscales show a stronger indication of what the scale is measuring. Subscale means for the aggregate sample for these scales were drive for thinness $M = 6.76$, ($SD = 6.80$), body dissatisfaction $M = 11.22$, ($SD = 8.24$), perfectionism $M = 10.93$, ($SD = 5.92$) and asceticism $M = 4.69$, ($SD = 3.37$).

Body mass index. Each dancer’s BMI was calculated by following the formula: kg/m$^2$, or weight divided by the square of height. The difference found between the BMI of recreational dancers ($M = 21.23$, $SD = 2.33$) and the BMI of preprofessional dancers ($M = 20.57$, $SD = 1.91$) was not significant. Between groups, 87% of the recreational studios’ participants were in the normal BMI range, with 4.3% of participants in the underweight and 4.3% of participants in the overweight range. In the preprofessional studios, 84.4% of participants were in the normal BMI range, with 15.6% of participants in the underweight and no participants in the overweight range. Overall, 1 out of 55 participants did not know her weight.

Body image silhouettes. In all cases, the discrepancy was in the direction of idealizing a smaller body image and a smaller ballet image than the perceived or the ideal image respectively. In the recreational sample, 56.5% reported a discrepancy between perceived versus ideal body image, compared to the 37.5% of participants from the preprofessional sample. Eighty-seven percent of recreational students reported a discrepancy between perceived versus the ballet ideal body, as did 87.5% of the preprofessional sample. For the comparison of ideal body image versus the ballet ideal body, 78.3% of recreational dancers and 62.5% of preprofessional dancers reported discrepancies.

Inferential Statistics

All tests were based on an alpha coefficient of .01, an adjustment helpful in decreasing the probability of making a Type I error, common after performing multiple inferential tests. Paired samples $t$ tests showed the differences in body image perception among the aggregate sample. Perceived body image ($M = 4.29$, $SD = .96$) was significantly larger, $t(54) = 4.98$, $p < .01$, than the ideal body image ($M = 3.74$, $SD = .82$). Perceived body image was also significantly larger, $t(54) = 10.35$, $p < .01$, than the ballet ideal ($M = 2.91$, $SD = .73$). A comparison of ideal body image and the ballet ideal demonstrated a significantly larger ideal body image than the ballet ideal, $t(54) = 9.40$, $p < .01$.

Between groups, the perceived body image of recreational dancers ($M = 4.61$, $SD = 1.03$) was significantly larger than the perceived body image of preprofessional dancers ($M = 4.06$, $SD = .84$). No significant differences were found between comparison groups on the ideal body image or the ballet ideal image.

Qualitative Analyses

To further understand the participants’ perceptions regarding body image relative to ballet, a number of qualitative questions were asked in the demographics survey. One item asked if the physical image of dancers should be changed. Of preprofessional dancers, 12.5% and of recreational dancers, 21.7% responded, “No, it is attainable for those who want it most.” When asked if she were striving to achieve this body type, 31.3% of preprofessional dancers and 30.4% of recreational dancers responded, “Yes.” Forty-one percent of preprofessional dancers and 52.2% of recreational dancers stated, “I think about striving to achieve this body type, but never act on these thoughts.” However, when asked if they would still strive or think about striving toward the ballet body if they were not ballet dancers, 68.8% of preprofessional dancers and 56.6% of recreational dancers affirmed, “No, I would like to be thin but maintain curves.” Specific questions were asked about how the dancers felt about their bodies, in a general sense and in ballet. The findings for the preprofessional sample are presented in Figure 1 and the findings for the recreational sample are presented in Figure 2.
Discussion

This study found no significant differences between the preprofessional and recreational dancers on BMI, ideal body image, ideal ballet body type, or across the four subscales of the EDI-3. However, recreational dancers had a heavier perceived body image than the preprofessional dancers. This is an important finding because there was not a significant difference when comparing the mean BMI scores for both samples. This discrepancy indicates a possible misperception in body image for recreational dancers. However, significant differences were found between perceived body image and ideal body image, perceived body image and the ballet ideal body type, and ideal body image and the ballet ideal body type in the aggregate sample.

The majority of dancers, both preprofessional and recreational, reported a desire to have a smaller body type than they currently have, and reported favoring an even smaller body type for ballet than their specified ideal body type. This finding is noteworthy because both groups endorsed a smaller ballet body ideal than their own nonballet body ideal. I hypothesized that the recreational dancers would not support this concept of having different ideals for their bodies in a general sense and specific to ballet as strongly as the preprofessional dancers. This hypothesis comes from the reality that weight is not commonly, if at all, discussed at recreational studios, thereby reducing dance environment pressures of conforming to the thin ideal. However, the results may indicate that the ballet subculture as a whole plays a larger role than the individual studio environments in promoting the ballet body ideal.

Data across the four EDI-3 subscales indicated an elevation in the eating disorder traits of the dancers. While a more definitive comparison could have been completed using mean data, the small sample size resulted in insufficient statistical power for these analyses.

FIGURE 1

Feelings about body: Pre-professional sample.

FIGURE 2

Feelings about body: Recreational sample.

Moreover, with scales that vary in ranges, it is difficult to compare the means between the two groups in the study. Relative to the normative sample, frequency data from the present study suggested that many of the dancers fell into potentially pathological categories on the Drive for Thinness and Perfectionism subscales.

Based on the responses to the qualitative questions, many dancers appear to have internalized the current ballet image and even to have thoughts about achieving it. In fact, 21.5% of recreational dancers said that the current ballet image should not be changed and is easily attainable, while only 12.5% of preprofessional dancers answered in the same manner. A possible explanation for this finding is that the preprofessional dancers may have experience in actually trying to achieve the ideal ballet body image through path of eating pathology and may have realized that, despite their efforts, not everyone can achieve this body type.

However, those who did support the ideal ballet image made it apparent that their support was confined within the ballet subculture and not in a general sense. The majority of the dancers sampled reported that they would not want this body type if they were not ballet dancers. Recreational dancers appeared less happy with their bodies, both with regard to the ballet subculture and larger society than the preprofessional dancers. Because there were no significant differences between BMI scores for the two groups, however, the attribution for this finding is either misperception of body image or a flaw of the BMI measurement. The collection of height and weight data was by self-report, the results of which could have been skewed by dishonesty or by misperception of their own weight. In addition, BMI calculations do not take into account muscle mass, which can skew the BMIs of the dancers as many tend to be muscular, which in reality increases their weight but not their BMI, or measure of body fat.
Limitations
Although the expected differences between preprofessional and recreational students in this study were not significant, there are some limitations to examine. Operational definitions for each type of studio were important in order to accurately categorize the participating studios as preprofessional or recreational. These definitions included most, but not all requirements of these particular types of studios. Furthermore, finding studios that met these definitions and agreed to participate was also difficult. Gaining permission from studio directors was troublesome; more elite preprofessional studios declined consent for participation in this study than did recreational studios. This may have resulted in skewed data.

Many questions used in the study were perception-oriented. Although these questions were useful, interpreting perceptual data is difficult. Furthermore, the scales in this study were self-report, which increases the chance of participants underreporting signs and symptoms of pathology. Although all participants were guaranteed complete anonymity, many young girls battling with secret eating problems might not trust a study such as this to remain anonymous. Perceptions of dancers likely have a bias, intentional or not. A study of nondancers for comparison would have been extremely helpful.

Using the BMI of the participants was helpful because the other instruments were self-report. Comparing the participants’ subjective thoughts to their calculated BMI (as per self-report of height and weight and the experimenter’s use of the BMI formula) was useful in exploring whether or not their perceptions were accurate (i.e. a participant feels that she is overweight even though her BMI indicates that she is underweight). Unfortunately, simply using one’s height and weight to determine one’s BMI leaves room for error, because it does not take into account bone density or muscle weight.

Future Directions
Despite the disadvantages of the present study, the findings could be useful in dance education. This study suggests that the ballet subculture may have a greater influence over all students than type of studio environment. Teaching the history of ballet, ballet aesthetics, and eating disorders and their consequences to all ballet students as a prevention (or in some cases, an intervention) technique may be extremely useful for those students who may only be learning this information through their peers. Young girls who are trying to gain acceptance into the ballet subculture tend to mimic the actions their peers model. Moreover, the source of this peer information may be questionable.

After observing the close similarities in characteristics shared among the pre-professional and recreational dancers, it would be interesting to see if girls with certain personality characteristics enter the art. Most dancers scored high on the scale of perfectionism on the EDI-3 and expressed a sense of competitiveness. A follow-up research question could center on whether ballet training has an influence on these student characteristics, or if students bring these characteristics with them into ballet class.

There are still some unanswered questions, which when answered, could help advance the knowledge of the ballet and body image relationship further. One unidentified factor is what exactly is the difference between preprofessional and recreational studios that affects students’ perceptions? The present study suggests that the influences of the two studios are not as different as once thought. Using recreational studio dancers for ballet research is still uncommon, so perhaps more studies will begin to examine this forgotten population of dancers.

References
Dementia is a significant public health issue that is rapidly increasing in prevalence. Hebert, Scherr, Bienias, Bennett, and Evans (2003) found that approximately 5 million Americans are affected by Alzheimer’s disease (AD), the most common type of dementia. Vascular dementia (VaD) is also increasing in prevalence and is currently the second most common form of dementia (The National Institute on Aging, 2007). This increase in prevalence is most likely due to the increasing lifespan of Americans due to improved access to and quality of health care services. Research estimates that approximately 2.4–4.5 million Americans currently live with AD (The National Institute on Aging, 2009). These statistics become even more significant when it is taken into consideration that a large proportion of the adult population, the so-called “baby boom” generation, is reaching the age at which there is a heightened risk for developing dementia. With this in mind, the Alzheimer’s Association (2009) estimates that nearly 16 million Americans will live with AD by the year 2050. The annual direct and indirect costs of dementia are estimated to be approximately $100 billion, with this figure expected to approach $118 billion by the year 2050 (The National Institute on Aging, 2007).

Functional impairment is a criterion for the diagnosis of dementia (American Psychiatric Association, 2000). The majority of research on functional decline relies on caregiver report (e.g., Guilmette, Temple, & Kennedy, 2008; Pratt, et al., 2007). Caregiver reports are attractive because they are considered to be valid and reliable, while also being brief and easy to administer (for further review, see Burns, Lawlor, & Craig, 2002 and Leifker, Patterson, Heaton, & Harvey, in press). In contrast, some previous research suggests that questionnaire-based assessments from informants, typically a family member or caregiver who has frequent contact with the patient, can be biased by insights, values, comparisons to previous states or other persons, and other events that may be occurring at the time that the report is given, thus making the accuracy of such reports debatable (Patterson, Goldman, McKibben, Hughes, & Jeste, 2001). Caregiver questionnaires are also susceptible to error from a number of other sources (Demers, Oremus, Perrault, Chapoulx, & Wolfson, 2000), including the nature of the relationship of the caregiver to the patient (Ready, Ott, & Grace, 2004) and perceived level of distress caused by caring for the patient (Clyburn, Stones, Hadjistavropoulos, &...
Tuokko, 2000). From a sampling perspective, if there is no caregiver present, then the patient is often excluded from studies of functional decline, possibly resulting in biased samples that might exclude single people and over-sample married couples.

Previous research examining questionnaire-based methods has given rise to the notion that such methods may not provide an accurate portrayal of the deficits that the patient is experiencing. Ready, Ott, and Grace (2004) compared the reports of informants to the self-reports of controls and patients with both AD and Mild Cognitive Impairment. These reports included questions addressing quality of life, depression, perceived level of severity, distress caused by symptoms, and clinical insight. The correlation between the informant report of memory abilities and actual, objectively-observed memory abilities was 0.62 for caregivers who lived with patients but only 0.18 for caregivers who did not live with patients. The difference between these two correlations was significant, indicating that caregiver reports can vary greatly depending on whether or not the caregiver lives with the patient.

Other studies also have raised questions about the validity of informant reports. Burke and colleagues (1998) observed that perceived level of burden and severity of patient symptoms strongly influenced the reports of informants. Another study concluded that age and physical/health status influenced the differences that were found between informant and self-reports of health related concerns associated with old age (Boyer, Novella, Morrone, Jolly, & Blanchard, 2004). Using the Assessment of Motor and Process Skills (AMPS; Doble, Fisk, & Rockwood, 1999), Boyer and colleagues found that family members were equally likely to overestimate functional impairment as they were to underestimate functional impairment. The researchers did conclude, however, that impairment was more likely to be overestimated in cases of mild rather than severe impairment. Similar findings from other researchers support the conclusion that informant report of functional impairment varies according to factors other than functional impairment itself (Zanetti, Geroldi, Frisoni, Bianchetti, & Trabucchi, 1999; Arguelles, Lowenstein, Eisdorfer, & Arguelles, 2001).

Performance-based assessment has the advantage of directly assessing functional capacities without concern for the biases of questionnaire-based methods. Performance-based assessment examines a participant’s ability to perform everyday activities (Instrumental Activities of Daily Living, or IADLs), such as cooking, financial management, and shopping. Furthermore, the use of performance-based assessment can be used to assess patients who do not have a close friend or family member to serve as an informant. Recent attempts have been made to develop reliable and valid instruments of everyday functioning, including the Texas Functional Living Scale (TFLS; Cullum et al., 2001), the Independent Living Scale (ILS; Loeb, 1996), and the Naturalistic Action Test (NAT; Giovannetti et al., 2007). Although each of these measures is still an indirect, clinic-based measure of everyday functioning, their validity is supported by data indicating that they are predictive of real-world functional indicators such as dementia status (Cullum et al., 2001), clinically-rated disability (Schwartz, Segal, Veramonti, Ferraro, & Buxbaum, 2002), unemployment (Heaton et al., 2004), and informant ratings of functional impairment (Cullum et al., 2001). Most of these instruments also have evidence for acceptable internal and test-retest reliability (Cullum et al., 2001; Loeb, 1996; Giovannetti et al., 2007). These instruments tend to be relatively easy to administer. However, a consistent weakness in many of these measures is that they do not comprehensively assess key domains of everyday functioning.

Heaton and colleagues (2004) devised a battery based on previously-published functional tests and newly-developed tests with the goal of assessing the key functional domains of cooking, financial management, medication management, shopping, and work skills, for use in a younger neurologically impaired population. The previously published instruments included the Direct Assessment of Functioning (Lowenstein et al., 1989), the Medication Management Test (Albert et al., 1999), and a standardized work sample (Valpar International Corporation, 1986, 1992). For the present study, we modified and supplemented the Heaton and colleagues functional impact assessment (FIA) for use in an elderly population. We added the Communications subtest of the Direct Assessment of Functioning because it is sensitive to functional decline in dementia (Lowenstein et al., 1989), and we eliminated the work sample because it is not relevant to these retired and severely disabled patients. The FIA is more comprehensive than previously-published batteries because it samples a wide range of behaviors: financial management, medication management, cooking skills, communication skills, and shopping ability.

Not only are most of the extant performance-based assessments limited in comprehensiveness, but the extant literature contains few studies exploring the relationship between questionnaire-based methods and performance-based assessment (e.g., Kempen, Stevernick, Ormell, & Deeg, 1996; Reuben, Siu, & Kimpau, 1992). The current study focuses on informant reports of functional deficits and seeks to explore the question of validity of such reports by comparing the scores obtained on questionnaire-based methods to performance-based assessment. In addition, we seek...
to compare the reports of both the patient and the informant to the FIA to determine the degree of inaccuracy of questionnaire methods by directly comparing them to performance-based assessment.

While previous studies (e.g., Evans et al., 2003) do indicate that performance-based assessment generates meaningful data, issues regarding the validity and reliability of performance-based assessment do exist. Miller and Linn (2000) argue that for a performance-based test to generate truly reliable data, a large number of tasks must be included in the test battery. Otherwise, the test might be biased to particular demographic groups. One of the advantages of the FIA over previous performance-based assessments (e.g., Albert et al., 1999) is the breadth of skills that are assessed.

We hypothesized that the FIA scores will be lower in the dementia group than in the control group. We also hypothesized that there will be a significant difference between the reports of the patients versus the informants in terms of functional ability. We expect that informant report will be more accurate than patient report when compared to performance-based assessment. We expect that the patients will be more likely than their informants to overestimate functional ability. That is, patients will be more likely than the informants to report normal functioning when impairment actually exists. For the purposes of this study, we assume that the FIA is a more accurate reflection of real-world functioning, although we recognize that this assumption is not directly addressed by the data in this study. Last, we expect that questionnaire methods will explain less than half of the variation in performance-based assessment.

**Method**

**Participants**
Participants included a mixed sample of dementia patients with either AD (n = 8) diagnosed according to NINCDS-ADRDA criteria or VaD (n = 8) diagnosed according to California ADDTC criteria. In addition, 16 demographically-matched control participants were recruited from both the Veteran’s Administration and from the community. Patients were diagnosed by a neurologist or a neuropsychologist in a clinical evaluation that was independent of this study. The Mini-Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975), which was administered as part of our test battery, mean score for the patient group was 20.38 (4.30), suggesting mild to moderate impairment.

Informants (n = 16) were defined as a family member (usually a spouse or child) or caregiver who spends at least 5 hours of contact per week with the patient. Patient and informant demographic information is provided in Table 1. Information regarding the nature of the relationship between the patient and the informant is provided in Table 2.

Participants were recruited from nearby Veteran’s Administration Memory Disorders and Neurology clinics, referrals, and from the community. If the patients were deemed too impaired to give informed consent, the legal representative of the patient consented on behalf of the patient. Patients were compensated $45 for the initial visit; informants were not compensated. All research was conducted at a state Veterans Administration Health Care System.

**Measures**

**Demographic questionnaire.** All participants completed a demographic questionnaire about relevant information such as age, education, and ethnicity in addition to questions pertaining to both individual medical and disease history and histories of his or her family.

**The Mini Mental Status Exam (MMSE).** To gauge the severity of dementia, the MMSE was administered.

**TABLE 1**

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Patients’ Informants</th>
<th>Controls</th>
<th>Controls’ Informants</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>16</td>
<td>16</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Age</td>
<td>80.38 (9.12)</td>
<td>65.63 (17.04)</td>
<td>76.73 (6.28)</td>
<td>65.29 (12.38)</td>
</tr>
<tr>
<td>Education</td>
<td>14.44 (3.10)</td>
<td>14.67 (2.02)</td>
<td>15.00 (2.30)</td>
<td>15.21 (2.64)</td>
</tr>
<tr>
<td>Ethnicity (% Caucasian/Non-Hispanic)</td>
<td>81.3%</td>
<td>68.8%</td>
<td>86.7%</td>
<td>92.9%</td>
</tr>
<tr>
<td>Sex (% Male)</td>
<td>93.8%</td>
<td>6.2%</td>
<td>80.0%</td>
<td>21.8%</td>
</tr>
</tbody>
</table>
to all participants. The MMSE is a widely used cognitive screening instrument that is sensitive to dementia. It has a maximum score of 30 (Folstein, Folstein, & McHugh, 1975).

The Dementia Deficits Scale (DDS). The Dementia Deficits Scale (Snow et al., 2004) informant and patient versions were administered to both the patients and controls and their informants. The DDS includes four subscales: meta-awareness (e.g., “Do you/your friend or family member think that you/your friend or family member have an illness affecting your memory?”); cognitive deficits (e.g., “Do you/your friend or family member ever get confused about where you/your friend or family member are?”); emotional deficits (e.g., “Do you/your friend or family member feel as though you/your friend or family member have emotional or psychological problems?”); behavioral deficits (e.g., “Are you/your friend or family member often more suspicious or less trusting than you/your friend or family member should be?”), and functional deficits (e.g., “Would you/your friend or family member be able to handle a small household emergency such as a plumbing leak or a small fire?”) There is a total of 35 questions on the DDS and questions are presented in a “yes” or “no” format. One point is given if the answer reflects the belief of impaired functioning. Higher scores on both the patient and informant versions indicate higher perceived levels of impairment.

Past research examining the reliability and validity of the DDS has concluded that the DDS is both reliable and valid. Furthermore, the DDS is favorable because it examines not just cognitive deficits, but emotional, functional, behavioral, and meta-cognitive deficits (Snow, et al., 2004). The DDS is a useful measure of awareness of one’s impairment because a patient’s report of problems can be compared to an informant report using the parallel patient and informant versions. In this study, we compared patient and informant reports with the percentage correct that the patient obtained on the FIA.

Procedures
Prior to the scheduled visit, informants were mailed a packet of questionnaires that included the informant version of the DDS for completion before the scheduled visit. The DDS is a widely used cognitive screening instrument that is sensitive to dementia. It has a maximum score of 30 (Folstein, Folstein, & McHugh, 1975).

The Function Impact Assessment (FIA).
Last, the patient was administered the FIA, which included tasks requiring the participant to balance a checkbook, make change, manage medications according to verbal and written instructions, dial a telephone, prepare a letter for mailing, pick groceries from a mock store from memory after both oral and written presentations of a shopping list, and prepare a mock meal of vegetables, bread, and pasta. Participants were rated according to standardized procedures on the basis of completeness of the task and preciseness of following the instructions. High scores on the FIA indicate lower levels of functional impairment. When scored, one point is given for proper completion of each of the major steps necessary to complete a given task. For example, when completing the financial management portion of the FIA, the participant is given one point for signing the check, one point for properly addressing the recipient of the check, one point for completing the numeric and written amount correctly, and one point for writing the date in the proper spot on the check. Previous studies of the FIA and its component subtests indicate that it is a reliable and valid measure of functional impairment in HIV infected persons (Heaton et al., 2004; Albert et al., 1999), and patients with dementia (Lowenstein et al., 1989).

Statistical Analyses
An independent samples t test was used to assess the differences between the controls and the patients on the percentage correct on the FIA. Because the assumption of homogeneity of variance was not met, a nonparametric Mann-Whitney U test was used to assess the degree of agreement between the patients and informants on the DDS. When seeking to determine the accuracy of self-versus informant-report, a Chi Square test was used within the patient group, comparing informant to patient group on presence or absence of functional impairment. Again, because the assumption of homogeneity of variance was not met, a nonparametric Spearman’s correlation was used to correlate the patient and informant reports with the percentage correct that the patient obtained on the FIA.

Results
Participant data
The patient and the control group were of similar age...
with similar levels of education (see Table 1). Because most of the patients were recruited at a VA medical center, most patients and controls were male veterans with female informants. The informants of the patients reported more contact hours per week than was the case with the controls, most likely due to fact that the patients were likely to require more care than the controls.

**Sensitivity of FIA to Dementia**

We hypothesized that the control group would display less functional impairment than the patient group as measured by the FIA. The current data supports this conclusion. In terms of the percentage correct on the FIA, the control group obtained a mean of 81.76 ($SD=8.44$) while the patient group obtained a mean of 45.00 ($SD=18.68$). Based on these means, a large effect size was found ($d=1.57$). An independent samples *t* test shows that the difference between the controls and the patients in terms of FIA performance was statistically significant, $t(28) = 6.75, p < .001$.

**Informant Versus Patient Self-Report**

We also hypothesized that there would be a significant difference between the reports of the patients versus the informants in terms of functional ability. As seen in Table 3, the informant and patient reports revealed significant differences for meta-cognitive deficits, cognitive deficits, functional deficits, and total score, with the informants reporting higher levels of impairment than the patients with dementia. There was no difference in reported emotional deficits.

#### TABLE 2

<table>
<thead>
<tr>
<th>Informant Characteristics</th>
<th>Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean duration of relationship (in years)</td>
<td>37.69 (21.51)</td>
<td>33.14 (26.32)</td>
</tr>
<tr>
<td>Mean number of contact hours per week</td>
<td>95.69 (72.42)</td>
<td>56.00 (61.27)</td>
</tr>
<tr>
<td>Nature of relationship</td>
<td>Spouse: 56.3% Child: 18.8% Friend: 12.5% Other relative: 6.3% Paid caregiver: 6.3%</td>
<td>Spouse: 50.0% Child: 14.3% Friend: 21.4% Other relative: 7.1%</td>
</tr>
</tbody>
</table>

#### TABLE 3

<table>
<thead>
<tr>
<th>Means, Standard Deviations, and U-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>DDS Meta-Cognitive Ability</td>
</tr>
<tr>
<td>DDS Cognitive Deficits</td>
</tr>
<tr>
<td>DDS Emotional Deficits</td>
</tr>
<tr>
<td>DDS Functional Deficits</td>
</tr>
<tr>
<td>DDS Total Score</td>
</tr>
</tbody>
</table>

*Note.* DDS = Dementia Deficit Scale. Higher score reflects greater impairment.

*Significant at the .05 level.

**Significant at the .01 level.
Accuracy of Informant Report

We also hypothesized that compared to patient performance on the FIA, informants would offer more accurate DDS reports than patients. In addition, we hypothesized that patients would be more likely than informants to overestimate functional ability as measured by the DDS. To address this issue, the mean percentage correct on the FIA for the control group was calculated (M = 81.76, SD = 8.44). One standard deviation below the mean of the control group was then established as the cut-off point for functional impairment. Using this standard, 14 out of the 16 dementia patients met the criteria for functional impairment. To establish informants who report functional impairment in the patient, a similar procedure was used for the questionnaire-based report of functional impairment. A cut-off score for the report of impairment was established using the mean DDS informant functional report for the control group (M = .21, SD = .58). Using these cut-offs, agreement or lack of agreement between informant report of functional impairment and FIA score was established. It was found that in three cases (18.75% of patient cases), the patient was found to be functionally impaired on the FIA, but the informant did not report functional impairment. In two cases (12.5% of patient cases), it was found that the informant reported functional deficits that were not supported by the FIA. In total, there was a lack of agreement between informant report and FIA percentage correct in 5 cases (31.25% of patient cases).

This procedure was then repeated to determine the agreement between patient self-report agreement and FIA score. In 8 cases (50% of patient cases), the patient underestimated his or her own level of impairment, meaning that his or her FIA score indicated impairment whereas the patient did not report impairment as defined by the cutoff score approach described above. Because the cutoff was close to the lowest score on the DDS, falling in the unimpaired range using the cutoff method on the DDS essentially means the patient was denying any functional impairment. In no cases did the patients over-report their impairment. The accuracy of informant- versus patient-reported impairment was nonsignificant, $\chi^2 (1, N = 15) = 3.27, p = .07$.

We used two nonparametric correlations (Spearman’s rho) to examine the relationship between the FIA percentage correct and the informant and the patient self-report functional DDS scores. We similarly examined the relationship between the FIA percentage correct and patient self-report (see Figure 1). While the correlations indicate that informant and patient reports of deficits were in the expected direction for the group as a whole—lower performance-based scores were associated with higher subjective reports of functional deficits—the correlations were only modest in magnitude. Neither informant report ($p = .41, p = .12$) nor patient report ($p = .47, p = .07$) reached statistical significance, but did indicate a trend toward significance.

Discussion

The differences shown between the FIA performances of the two groups indicated that the FIA is an instrument that is sensitive to the presence or absence of dementia. Furthermore, correlations between the FIA and self- and informant-report of functional impairment were not significant, although the direction of the association was in the expected direction. The central hypothesis of this study, that informant report would explain less than half of the variance in performance-based assessment was supported by the finding that the correlation between informant reports and the performance-based measure were less than .5 ($r = .41$ for informant report, $r = .47$ for patient report, neither of which were statistically significant). If one assumes that the performance-based measure is a valid measure of everyday functioning, then one can conclude that questionnaire-based approaches may either be inaccurate estimates of real-world functioning or may estimate a different element of functioning than the FIA. As noted before, there is evidence from multiple studies that our performance-based measure is a valid predictor of real-world functioning, although it must be acknowledged that the FIA has not been validated.
with direct observation of patients’ completion of daily tasks. Therefore, our data provide only indirect support for the conclusion that questionnaires are imperfect measures of real-world functioning. Considering the practical problems with observing patients’ actual daily functioning in their homes, and the fact direct observation is rarely performed in research or clinical arenas, our data reflect the best possible estimates given such limitations.

The data also show that almost one third of the informants are either overestimating or underestimating functional impairment, while the patients are strictly underestimating impairment. The informant and patient reports were discrepant for meta-cognitive deficits, cognitive deficits, functional deficits, and total score with the informants reporting higher levels of impairment than the patients with dementia. There was no difference in reported emotional deficits, probably reflecting the low rate of self- and informant-reports of emotional dyscontrol in this sample. Because dementia affects one’s cognitive abilities, a plausible conjecture may be that dementia patients are not reporting accurately due to the fact that the patients have deficits in terms of meta-cognitive skills. It may be the case that those with dementia are simply too impaired to perceive their own deficits. Because of this possible lack of self-awareness and ability to self-assess, direct assessment of functional capacities or informant report should be used when assessing a patient’s functional and cognitive status. It is worth noting that about two thirds of the informants were accurate in their report of impaired patient functioning. It could be useful to determine if certain informant characteristics could predict whether or not they are accurate. It would also be useful to employ a questionnaire with a wider range than the DDS cognitive scale so that degree of accuracy of the informant report could be more reliably assessed.

The FIA is limited by the fact that the test is administered in a laboratory setting and may therefore not be an accurate reflection of how the patient functions within his or her home. Another possible limitation of the current investigation is that the sample size was small, thus limiting the power. However, this small sample size did not prevent statistically significant findings between patient and control groups. It is possible that a larger sample would result in the correlations between questionnaire methods and performance-based measures being statistically significant, but a larger sample size would probably not change the modest magnitude of the correlation.

Future studies should employ a more demographically diverse sample (especially with more women than were obtained by this primarily military veteran sample) to improve generalizability of these findings. Future consideration should also be given to establishing a test that assesses the participant within his or her home.

Further research is also necessary to examine the test-retest reliability of this new performance-based functional battery, so that such a battery can be incorporated into longitudinal studies of disease progression and treatment. It will also be interesting to use the FIA to further build on past research that investigates the role of patient and informant depression, and perceived level of caregiver distress and how it affects the validity of reported symptoms, as some of the previously-mentioned studies have noted. Larger sample sizes will be needed to examine the effect of informant demographic variables such as age, level of education, and gender, in addition to other variables of interest such as duration of relationship with the patient, length of time spent caring for the patient, and severity of the patients’ symptoms, on the reliability of informant reports. Future research may also focus on these variables and their effect on the accuracy of reporting functional impairment.

Past research looking at varying topics and populations within clinical psychology suggest that use of multiple methods and data collection techniques is optimal because it allows researchers to rule out possible alternative explanations to findings while also examining research questions that may not otherwise be possible when using a single method or data collection technique (e.g., Anastasi, 1997). Our findings provide support for the notion that studies of functional impairment are likely to have better accuracy by including a performance-based measure of everyday functioning when used in conjunction with other methods such as questionnaires and/or neuropsychological testing. Especially in studies of disease progression and treatment effects, reliance solely on informant reports may lead to incorrect estimates in change over time. As the number of people with cognitive disorders of aging increases, the need for accurate estimates of disease burden is a significant public health issue. Improved methods of estimating functional impairment could lead to better planning for future health care needs and more accurate estimates of the effectiveness of treatments.

References


Prentice Hall.


In recent years, much research has been conducted in the interdisciplinary field of second language acquisition (SLA) as a result of expanding interest by linguists, psychologists, neuroscientists, and educators. Two related questions relevant to the aims of SLA research are whether instruction can facilitate SLA, and if so, which type of instruction is most beneficial. Although these questions may seem most relevant to applied linguistics, it is likely that their answers can be found in the basic science of the neurocognitive mechanisms that allow humans to acquire a second language (L2). The present research seeks to address both of these issues through the means of an experimental study of L2 vocabulary learning in which the effectiveness of two prominent instructional methods is compared to non-instructed, incidental learning, and in which individual differences in neurocognitive processing are examined during SLA.

The study examines which type of instruction facilitates SLA and whether prior knowledge of an L2 expedites subsequent re-learning, highlighting the importance of the cognitive processes subserving SLA.

Many approaches to L2 instruction have arisen during the past century, but two of the most well-known are the grammar-translation approach and the communicative approach. The grammar-translation approach, the older and more established of the two methods, originated from practices used to teach Latin to adolescent students enrolled in boarding schools in the 19th and early 20th centuries, and it remains entrenched in many L2 classrooms and curricula—especially at the novice level—to this day. Omaggio-Hadley (1993) points out that this approach is rooted in the belief that an L2 cannot be comprehended until the syntactic framework of the L2 has been acquired. Thus, in the grammar-translation approach, instruction is focused on the grammatical structure of the L2, and specific emphasis is placed on the accuracy of L2 spoken language.

A mini-lesson in Spanish vocabulary was taught to participants unfamiliar with the language using stimuli representative of one of two instructional methodologies, the grammar-translation approach or the communicative approach. Participants taught using the communicative approach recalled significantly more Spanish vocabulary words than those assigned to an uninstructed control group on a short-term post-test, but no significant differences emerged between any other groups on this measure. No differences in recall were found between any of the groups on a long-term post-test. Participants demonstrating some prior knowledge of Spanish recalled a significantly greater number of vocabulary words on the short-term post-test than participants unfamiliar with Spanish. The results suggest that instruction and prior knowledge of a second language can facilitate the early stages of second language acquisition.

Laura M. Morett
Washington College

The Effect of Instructional Method on Second Language Acquisition: An Examination of Some Contributing Factors

Laura Morett is now a PhD student of cognitive psychology at the University of California, Santa Cruz. This research was conducted as a Senior Capstone Project for the Department of Psychology of Washington College.

I thank Lauren Littlefield for serving as the advisor to this research. Additionally, I thank three anonymous reviewers as well as the editor of this journal for providing helpful feedback on earlier versions of this paper.

Correspondence concerning this paper should be addressed to Laura Morett, Department of Psychology, University of California, Santa Cruz, Santa Cruz, California, 95064. Email may be sent to lmorett@ucsc.edu.
and written output. In general, practitioners of the grammar-translation approach believe that if students are taught the rules governing the grammar of the L2, proficiency in the L2 will follow of its own accord.

Some research suggests that the teaching of L2 grammar is an essential element of structured L2 instruction, especially that designed for adolescents and adults. In a meta-analysis of the effect of instruction on SLA, Norris and Ortega (2000) reported average effect sizes of 1.0 and 0.93 for studies in which students were taught some aspect of the target language, but only a 0.54 effect size for studies in which grammar was not explicitly taught. Thus, the results of this meta-analysis provided systematic quantitative evidence in favor of grammar instruction for SLA. More recent evidence in favor of instructional methods similar to the grammar-translation approach can be found in Klapper and Rees (2003), who tracked the L2 knowledge of two groups of ability-matched college students taught using a methodology based either on form-focused instruction, in which grammar is taught explicitly, or communicative skills, in which emphasis is placed on conveying meaning in the target language at the expense of grammatical correctness. These authors found that the students taught using the form-focused method acquired the L2 more quickly and accurately than their peers who were taught using the communicative approach, as measured by periodic standardized assessments of L2 grammar and proficiency.

Laufer (2006) compared the acquisition of novel L2 vocabulary by high school students who had obtained an advanced level of proficiency in the L2 given either a text comprehension and discussion exercise in which the vocabulary words appeared with clues to their meanings or a vocabulary learning exercise in which vocabulary words were presented with their meanings and examples of their usage. The results of this study revealed that, when students were tested on the meanings and uses of the vocabulary words immediately following the activities, those who had completed the vocabulary learning exercise learned more of the vocabulary words than those who had completed the text comprehension exercise, providing evidence that a focus on form facilitates short-term L2 vocabulary learning. Taken together, the results of these studies seem to suggest that adolescent and adult students are able to more effectively acquire various aspects of an L2 when taught using methods similar to the grammar-translation approach, highlighting the usefulness of instruction focusing on the grammatical structure of the target language.

The communicative approach to L2 instruction arose as an alternative to instructional methods emphasizing grammar acquisition. In contrast to grammar-focused methods, which were based on the tenets of behaviorism, emphasizing rote learning of the structure of the target language, the communicative approach focuses on the development of communicative competence—that is, learners’ ability to communicate in true-to-life situations. There is disagreement about whether any grammar should be explicitly taught within the context of the communicative method; some scholars, such as Krashen (1982), believe that L2 grammar is learned implicitly as long as students are provided with comprehensible input in the L2, whereas other scholars, such as Long (2007), believe that it is beneficial to draw students’ attention to certain grammatical structures within the context of a task-based lesson in the L2. In contrast with the developmental sequence of SLA hypothesized within the grammar-translation model, the communicative approach posits that proficiency in the L2 precedes explicit knowledge of L2 syntax, highlighting the need for meaningful, task-based activities within L2 instruction.

There is a body of research that supports the communicative approach and related teaching methodologies that do not focus explicitly on L2 grammar. One prominent example is Van Patten and Cadierno’s (1993) experimental study supporting the efficacy of a communicative approach to L2 instruction. In this study, students were taught various object pronouns in Spanish using either the Processing Instruction approach, in which students completed a variety of meaningful comprehension exercises composed of sentences including an object pronoun, or a Traditional Instruction approach, in which the functions of the object pronouns were described to students, and students then practiced producing them within the context of traditional structured drills. Performance on sentence-level comprehension and production tests revealed that students taught using the Processing Instruction approach interpreted and produced object pronouns in the L2 more accurately than students taught using the Traditional Instruction method, thus demonstrating a deeper understanding of the pragmatic uses of object pronouns. The results of this study were replicated using similar assessments of L2 proficiency for several syntactic structures commonly found in Spanish, including the preterit (immediate past tense) (Cadierno, 1995), the ser/estar distinction (Cheng, 1995), and the subjunctive tense (Farley, 2001), as well as in other languages such as French (Van Patten & Wong, 2004) and Italian (Benati, 2001). Additionally, Skala (2003) found that instructional methods theoretically related to the communicative approach were more effective for teaching high school students of beginning French than grammar-focused methods, providing further evidence supporting the efficacy of communicative
approaches to SLA. All of this research suggests that, in contrast to the research supporting the grammar-translation approach, it is unnecessary to explicitly and extensively provide instruction about the structure of the L2, because as long as students are provided with comprehensible input in the L2, they will acquire the grammar of the L2 implicitly.

In order to fully understand the process of L2 acquisition, it is necessary to consider the neurocognitive mechanisms by which L2s are learned in addition to the effects of instruction on L2 acquisition. One cognitive process that is essential to L2 acquisition is working memory. It is generally agreed upon that working memory is an essential component of many higher-order cognitive processes, such as SLA. One theoretical construct of working memory relevant to SLA is the phonological loop, a system within working memory that continuously repeats the sounds of words that one has heard or read (Baddeley & Hitch, 1974). The phonological loop is hypothesized by some researchers to play an especially important role in the acquisition of novel words (Baddeley, Gathercole, & Papagno, 1998). Research on the relationship between the phonological loop and SLA has revealed that the accuracy of the repetition of pseudo-words (e.g., mokka) can predict the acquisition of L2 vocabulary (Ellis, 1996; Service, 1992). One piece of empirical evidence for the connection between working memory and SLA comes from correlational studies in which researchers observed that the ability of British schoolchildren to repeat unfamiliar pseudo-words contrived in accordance with the phonotactic rules of the French language was related to their knowledge of French vocabulary (Gathercole & Thorn, 1997; Thorn & Gathercole, 1999).

Furthermore, the modality in which verbal stimuli are presented can influence the functioning of working memory in language learning and processing. In discussions of articulatory suppression (that is, interference in the functioning of the phonological loop by words presented simultaneously with words to which one is attending), Baddeley, Thomson, and Buchanan (1975) claimed that the processing of purely aural or visual verbal stimuli is deleterious to the functioning of working memory. They pointed out that when verbal information is presented only in the aural modality, the complementary visual code is absent, and vice versa, resulting in impoverished information processing and representations. These assertions support the notion that stimuli rich in multimodal content are needed to ensure optimal functioning of working memory in language acquisition and processing.

Laterality, the differential activation of the cerebral hemispheres, is a neurological construct that also affects L2 acquisition. In general, research on the role of laterality in information processing has shown that the left frontal cortex encodes and processes verbal stimuli, whereas the right frontal cortex encodes and processes visuospatial stimuli. In some studies (e.g., Kelley et al., 1998), bilateral hemispheric activation has been observed during the processing of stimuli containing both verbal and visual materials. In accordance with this traditional view of hemispheric specialization, some researchers (Galloway & Scarcella, 1979; Galloway, 1981) have found little evidence for the involvement of the right hemisphere in the processing of an L2; however, other researchers (Wesche & Schneidermann, 1982) have reported some evidence for the involvement of the right hemisphere in the beginning stages of SLA. This latter finding is consistent with Baddeley’s (2003) hypothesis of increased involvement of the right hemisphere in language processing given a task that involves a high cognitive load (such as initial-stage language learning), as well as with Sundermeier, Virtue, Marsolek, and Van den Broek’s (2005) observation of increased activation of the right hemisphere when processing unfamiliar verbal information. These findings suggest that it may be beneficial for people to learn an L2 in an instructional setting using multimodal materials that draw upon the processing specializations of both hemispheres, providing them with multiple channels by which they can establish representations of the L2 and later draw upon those representations to comprehend and produce the L2.

The purpose of this experiment was to compare the effectiveness of two prevalent L2 instructional methods, the grammar-translation approach and the communicative approach, in early-stage SLA. To that end, two experimental conditions were created in which materials representative of each method were used to teach L2 vocabulary. The first, the verbal condition, included purely verbal stimuli in the form of L2 vocabulary words paired with their translations in the participants’ native language, in accordance with the use of translations within the grammar-translation approach. The second, the integrated condition, included a combination of verbal and visual stimuli, in the form of vocabulary words paired with illustrations that symbolized their meanings, in accordance with the use of non-linguistic representations within the communicative approach to convey the meaning of L2 utterances. In order to examine the role of the cognitive and neurological processes involved in SLA, participants’ working memory efficiency and cerebral hemispheric activation were measured. Based on the findings of research highlighting the importance of the phonological loop in L2 acquisition (e.g., Thorn & Gathercole, 1999), I predicted that participants with more efficient working memory capacity would learn a
greater number of L2 vocabulary words. With regard to laterality, given the results of studies suggesting that visuospatial and verbal stimuli are processed differently in each hemisphere and that both hemispheres are active when both types of material are processed (e.g., Kelley et al., 1998), I predicted that L2 vocabulary would be more effectively learned under the integrated condition due to its use of both visuospatial and verbal materials.

**Method**

**Participants**

All participants were undergraduate students at a small liberal arts college located on the East Coast of the US, aged 18-23 years ($M = 19.09; SD = 1.16$). Thirty-two participants (6 men, 26 women), all of whom claimed to be unfamiliar with Spanish, the L2 presented in the experiment, volunteered in partial fulfillment of a pre-experimental L2 vocabulary quiz, which were interpreted as evidence that the participants may have had some residual knowledge of Spanish.

**Materials and Apparatus**

Videos used in this study featured the author teaching a lesson in Spanish vocabulary that consisted of 30 words from the introductory text *Puntos de Partida* (Knorre, Dorwick, Perez-Girones, Glass, & Villareal, 2004). The version of the video shown in the verbal condition included words paired with their translations, and the version shown in the integrated condition included words paired with pictures that illustrated their meanings. The duration of the videos was approximately 7 min each.

Forms used in this experiment included a demographic survey, a vocabulary list, a vocabulary worksheet, a pre-test, a metacognitive survey, and a short-term post-test and long-term post-test. The demographic questionnaire was used to measure several general factors of interest, including sex, age, academic background, and prior experience in SLA. The pre-test, which was intended to screen participants for prior knowledge of the L2, consisted of 5 Spanish words, chosen at random from the vocabulary used in the teaching portion of the experiment, to be translated into English. A vocabulary list and worksheet containing the words and their referents presented in the videos allowed participants to facilitate their L2 learning during and immediately after the lesson. The short-term and long-term post-tests included all of the words presented in the video, with half of the questions presented in a visual format and half presented in a verbal format, counterbalanced across test versions. Subjective and affective factors, such as enjoyment of the experimental lesson and perception of the effectiveness of the instructional method, were measured using a 7-question metacognitive survey.

The software package *Laboratory in Cognition and Perception, Third Edition* (Levy & Ransdell, 1998) was used to administer a computerized version of the Sternberg task (1966), a standardized task of working memory efficiency in which a series of digits varying in length is presented, followed by a special symbol signifying recall, and then by a single digit to which participants respond as quickly and accurately as possible, indicating whether it was present in the previously-presented set of digits.

In the neurological portion of the experiment, the Edinburgh Inventory (Oldfield, 1971) was used to determine handedness of participants, given that the results of past research (e.g., Knecht et al., 2000) have revealed a relationship between handedness and hemispheric dominance in language processing. Hemispheric activation was measured using a transcranial Doppler, an instrument that measures blood flow through the middle cerebral arteries of each hemisphere via ultrasound technology. Readings were averaged over a two-minute period during the experimental task for analysis of hemispheric activation during the online processing of target language vocabulary.

**Procedures**

**Main session.** Based on experimental session, participants were assigned to one of the three conditions described above: verbal, integrated, or control. Seven participants (3 men, 4 women), all right-handed native speakers of English, took part in the neurological component of the experiment. In addition to an independent analysis of these participants’ hemispheric cerebral hemodynamics, their results on the behavioral measures were included in the overall analysis. Small groups of 3-6 participants were recruited for each experimental session using a sign-up sheet with multiple lines for each time slot. To ensure that all participants had at most minimal knowledge of Spanish, the L2, it was stipulated on the sign-up sheet that volunteers must have had no more than one academic year of Spanish instruction in middle or high school or no more than one semester of Spanish instruction in college. When participants arrived for the experimental sessions, after providing their informed consent, they completed the pre-test, which was immediately scored by the experimenter to ensure minimal prior knowledge of L2 vocabulary. Participants then proceeded to complete
the demographic survey and then the Sternberg task of working memory efficiency.

In the learning portion of the experiment, participants in the experimental conditions were seated in an arrangement similar to that of a typical classroom. Prior to the lesson, participants were given a printed vocabulary list containing the vocabulary words paired either with their translation or a pictorial representation of their meaning in accordance with the stimuli presented in the instructional segment of the video for their assigned condition, and were encouraged to use the list to follow along with the video. After the lesson, they completed a non-evaluative review sheet containing all of the vocabulary words, half of which were represented by purely verbal stimuli (i.e., translations) and half of which were represented by visual stimuli (i.e., images). The experimenter then played the second part of the video, in which the correct answers were provided for the exercises on the review sheet. After watching the second part of the video, participants filled out the metacognitive survey, completed the short-term post-test, and were debriefed, ending the experimental session.

The procedure used for the control group was identical to that used for the experimental groups, except that the videos containing the experimental lessons were not played and the vocabulary lists, practice worksheets, and metacognitive surveys were not distributed. The purpose of the control condition was to determine the guessability of the questions on the post-tests, making it possible to account for the effects of random guessing on the results for participants assigned to the experimental conditions. None of the sessions for the experimental or control conditions lasted longer than 60 min.

Long-term post-test session. A subset of 18 participants (4 verbal, 8 integrated, 6 control) from the experimental and control conditions returned 2 weeks after the main session to complete a follow-up session, for which they received an additional experimental credit. In this session, participants provided their informed consent a second time, took a long-term post-test that was similar but not identical to the short-term post-test, and were then debriefed and released. None of these sessions lasted longer than 15 min.

Neurological component. Participants who volunteered for the neurological portion of the study performed all tasks in the same sequence as participants in the experimental groups of the behavioral component of the study. The only differences in procedure were: (a) participants were run individually; (b) participants were assigned either to the verbal or integrated condition, but not to the control condition; and (c) blood flow velocity through the middle cerebral arteries was measured in mm/s via Transcranial Doppler during the instructional segment of the video, as participants learned L2 vocabulary. After completing all tasks, participants were debriefed and released. None of the sessions in the neurological component of the study lasted longer than 75 min.

Results

A univariate analysis of variance (ANOVA) procedure confirmed the presence of a significant difference in the short-term post-test scores of participants assigned to the control, verbal, and integrated conditions, F(2, 29) = 4.34, p = .02; partial η² = .23 (see Figure 1). An examination of Bonferroni-corrected post-hoc analyses revealed that the short-term post-test scores of the integrated and control groups differed significantly, p = .02, but that the scores of the verbal and control groups (p = .23) and the integrated and verbal groups (p = 1.0) did not differ substantially. Among participants assigned to the experimental conditions, significantly higher scores on the short-term post-test were achieved by those who obtained a score of 1/5 correct on the pre-test than those who scored 0/5 correct, t(22) = 22.69, p < .001. There were no differences between the scores of participants assigned to the experimental and control conditions on the long-term post-test.

Correlational analysis revealed a non-significant positive relationship between working memory efficiency and vocabulary learning, r(33) = .33, p = .07. The mean for participants’ reaction time during the Sternberg task, a measure of working memory efficiency, was 4644.03 ms (SD = 1093.36 ms), and the mean for scores on the short-term post-test, which measured vocabulary understanding, was 18.03 (SD = 7.75)
knowledge, was 18.91 items correct (SD = 4.50). These statistics suggest that participants with lower reaction times on the Sternberg task, which measures working memory efficiency, tended to obtain higher scores on the short-term post-test by recalling more L2 vocabulary words, resulting in the observed positive correlation between working memory efficiency and vocabulary retention. No significant differences in hemispheric activation during vocabulary learning were observed between participants in the verbal and integrated conditions, as measured by the velocity of blood flow to the left and right cerebral hemispheres via the middle cerebral arteries.

**Discussion**

The results of this study demonstrate a significant advantage of instruction and prior knowledge on the acquisition of L2 vocabulary by novice adult learners. A significant advantage of assignment to the integrated condition, in which instruction in the L2 was delivered using materials that included both verbal and visual stimuli, over assignment to the control condition, in which instruction in the L2 was not provided, was observed in the short-term post-test scores of participants. However, no significant difference was observed between the short-term post-test scores of participants assigned to the verbal and integrated conditions, suggesting that neither instructional method is more effective than the other. The difference between the short-term post-test scores of participants who scored a 1/5 and a 0/5 on the pre-test indicates that prior knowledge of an L2 may be advantageous when (re)learning vocabulary from the same L2, supporting the memory savings hypothesis of Ebbinghaus (1913). Taken together, these results highlight the role of instruction and prior knowledge of the L2 on early-stage SLA, demonstrating the importance of taking advantage of one’s existing cognitive processes in SLA.

The significant difference in L2 vocabulary acquisition observed between participants assigned to the integrated condition and the control condition reflected in the short-term post-test suggests that the communicative method may be an effective way to jumpstart initial SLA. The benefit of a communicative, integrated approach in initiating the process of SLA is consistent with the findings of Van Patten and Cadierno (1993) and Skala (2003), which demonstrated the effectiveness of task-based, meaning-focused activities that make use of multimodal stimuli in facilitating the acquisition of vocabulary and grammatical competence in beginning- and intermediate-level L2 learners. In both these past studies as well as the integrated condition of the study described here, students were encouraged to associate L2 forms with their meanings, thus facilitating their access to the mental lexicon, in which linguistic forms are stored and associated with their referents. Nevertheless, although the results of this study suggest that the communicative approach may convey a significant immediate advantage over incidental learning of the L2, the restriction of the advantage to the short-term post-test coupled with the lack of a significant difference between the verbal and the integrated conditions in either the short-term or the long-term post-tests suggests that it is premature to definitively conclude that the communicative approach is an overall more effective method of L2 instruction than the grammar-translation approach.

The lack of a significant difference observed between the short-term post-test scores of participants assigned to the verbal and integrated conditions could be due to a number of factors. One possibility is that it may be attributable to the similarity of the two conditions, as manifested in the L2 vocabulary words, which were presented as text, and the format (L2 word + translation/visual representation) of the stimuli in both conditions. If this is the case, differences in L2 vocabulary retention may only become noticeable when the vocabulary words are presented in a single modality and/or in different formats. However, aside from the difficulty of creating purely visual stimuli to convey linguistic information, prior research has shown that impoverished verbal stimuli can negate the normal processing of the phonological loop (Baddeley, Thompson, & Buchanan, 1975), so the feasibility of creating such stimuli is uncertain. It may also be the case that a longer period of exposure to the L2 is necessary for any differential effects due to instructional methodology and stimulus type to emerge. The research of Skala (2003) and Klapper and Rees (2003) suggest that this possibility may be valid, given that the findings were based on observations made over the course of a semester and four years, respectively. At any rate, longitudinal research on the role of instructional method in SLA should be conducted so that a more concrete conclusion can be reached.

The greater number of L2 vocabulary words correctly associated with their meanings by participants who scored 1/5 correct on the pre-test as opposed to those who scored 0/5 correct demonstrates the importance of savings in relearning an L2. According to Ebbinghaus (1913), prior exposure to information allows for faster learning in subsequent trials, and the increase in efficiency between trials is referred to as savings. The significance of the advantage for participants who scored 1/5 items correct on the pre-test suggests that savings due to prior learning of the L2 may be quite substantial. When considered in combination, the benefit of prior knowledge of the L2 and the non-
significant trend toward a positive correlation between L2 vocabulary learning and working memory efficiency suggest that the episodic buffer, a component of working memory proposed by Baddeley (2000) to package information from working memory for storage in and later retrieval from long-term memory, may be one of the cognitive mechanisms that subserves SLA. According to Baddeley, the episodic buffer channels information from working memory into a unified, multi-faceted code for long-term storage and later retrieval. If working memory efficiency and prior exposure to the L2 facilitate vocabulary learning as a result of the episodic buffer’s ability to transfer information between working memory and long-term memory, it is possible that the episodic buffer may also play a role in long-term SLA, suggesting that domain-general cognitive mechanisms play a key role in supporting SLA. Although these effects might not be manifested as strongly in long-term SLA as in early-stage L2 vocabulary learning, it is possible that they affect SLA in subtle ways, possibly contributing to individual learning style preferences. Further research should attempt to gauge the long-term impact of prior knowledge of the L2 through a longitudinal design, helping to further clarify the roles of long-term memory savings and working memory in SLA.

Measures of cerebral blood flow velocity gained using TCD showed no significant differences in activation between the two hemispheres during L2 vocabulary learning under the two conditions. This finding is interesting in light of the consistent findings from numerous past experiments that the left hemisphere is more active in the processing of verbal stimuli (see Jung-Beeman & Chiarello, 1998, for a review). The lack of laterality differences may be an indication that participants used mental imagery to process the purely verbal stimuli presented in the verbal condition, resulting in increased blood flow to the right hemisphere under this condition, which may have disguised any differences in hemispheric blood flow due to condition. It is also possible that the observed effects could be attributed to a lack of sufficient data due to the small size of the participant pool for the neurological component of the study. Whatever the case, further research into the role of laterality in the initial stages of SLA should be conducted to clarify these results.

Overall, consistent with prior research (see Norris & Ortega, 2000), the results of this experiment provide evidence for the short-term benefits of instruction and prior knowledge on SLA. More specifically, the results demonstrate that although materials similar to those used in the communicative method may provide L2 learners with an initial benefit over incidental learning in vocabulary acquisition, this benefit does not extend to long-term retention. Perhaps the findings of this study will be informative to foreign language teachers, applied linguists, and cognitive psychologists alike, helping them to better understand the relationship between the roles of instruction and the neurocognitive mechanisms underlying SLA.

References


Approximately 3 out of every 10 older adults will experience some type of unexpected fall in a calendar year. Of those, 5-15% will suffer a serious injury that includes lacerations, bone fractures, serious soft tissue damage, or head trauma. These injuries can lead to a decline in quality of life, inability to successfully age in place, and erosion of financial security (Cook, Yearns, & Martin, 2005; Tinetti, Gordon, Sogolow, Lapin, & Bradley, 2006; Tinetti, Richman, & Powell, 1990). Although the proportion of elderly persons over 65 experiencing a fall is approximately 30%, the ratio dramatically increases to 50% for those persons over 80. Consequently, the sixth leading cause of death for persons aged 65 and over is unintentional injury, most often resulting from a fall. Public health costs for acute care and services for older persons who have fallen were estimated to be $19.5 billion in 2000 (Finkelstein, Fiebelkorn, Corso, & Binder, 2004; Sattin, 1992; Tinetti et al., 2006). Given the individual and societal costs of accidental falls, prevention efforts are critical and merit attention by researchers (Finkelstein et al., 2004).

Falls and fall prevention have been studied extensively using a variety of interventions (Jørstad, Hauer, Becker, & Lamb, 2005; Finkelstein et al., 2004; Sattin, 1992). These actions have included, but are not limited to, free installation of grab bars and other safety devices for the occupant, balance training, and correction of lighting deficiencies (Gitlin et al., 2006; Stevens, Holman, & Bennett, 2001). The approaches, while successful, are not cost effective for use in reducing falls in the general population (Tinetti et al., 2006). Other approaches to fall prevention have implemented exercise programs conducted in senior centers to improve balance, gait, and self-efficacy. These too have displayed some measure of success. However, due to transportation challenges, not every eligible participant is capable of accessing an exercise program located outside the home (Li, McAuley, Harmer, Duncan, & Chaumeton, 2001).

Less expensive interventions involve distribution of educational materials focused on reducing the risks associated with falls. In some cases, informational materials are discussed and given to patients at doctor appointments; in other settings, patients are exposed to educational information via posters and brochures displayed in doctors’ waiting rooms. Interventions conducted in this manner have yielded some success, but do not reach those seniors without a primary care

---

**Improving Bathroom Safety Among Community Dwelling Older Adults**

We randomly assigned older adults to receive mailed educational materials about prevention of falls in the bathroom (n = 87), or no materials (n = 89). After 4 weeks, we contacted participants by telephone to assess self-reported change and intention to install bathroom safety features, as well as self-efficacy for maintaining physical balance. Compared to control participants, intervention participants were more likely to report making actual bathroom safety improvements, intent to take action, and serious intent to add safety features. These intentions were also related to self-efficacy for maintaining balance. Safety intervention materials can be tailored for specific populations while at the same time being fiscally sensitive.

**Julia L. Murphy**
**Ann M. Steffen**
University of Missouri—St. Louis

---

* Faculty mentor
Several studies have employed mailed educational information and have met with mixed success (Sevick et al., 2007). Using mailed educational items and verbal instructions, Sevick et al. reached 239 voluntary participants and assessed change by a survey returned in the mail. There were no statistically significant differences in behavioral change for those who received verbal instructions compared to those receiving printed materials. In another study, Solomon et al. (2005) mailed educational materials about fall risk reduction to participants over 65 years. No intervention impact was evident after an 8-month follow-up period; however, because the outcome assessment procedure resulted in a modest 54% response rate, the true effect of the intervention is unclear.

Most of the existing literature in fall prevention targets multiple areas of the home, such as kitchens, hallways, interior/exterior stairways, bedrooms, living rooms, and bathrooms (Gitlin et al., 2006; Stevens et al., 2001). Poor intervention results could be due to information overload; when faced with a large number of suggested household renovations, none are undertaken (Cook et al., 2005). In contrast, Heslin et al. (1992) documented an effective fall prevention strategy that focused solely on a single type of fall (e.g., falling out of bed) for residents living in an institutional setting. By emphasizing a single, highly vulnerable area for falls, researchers can increase the likelihood that participants engage in behavior change. Because the bathroom is the highest risk area for falling by community dwelling seniors, targeting this space for a prevention intervention makes sense.

When selecting a theoretical model to guide fall prevention efforts and outcome assessment, it is important to choose a model that conceptualizes change on a continuum, moving from levels of intentionality to actual behavior change. Prochaska’s Transtheoretical Model of Change (Prochaska & DiClemente, 1982) is a useful framework for this purpose, and describes five stages: (a) precontemplation with no intention of changing; (b) contemplation with intention to initiate change within the next 6 months; (c) preparation with intention to make a behavioral change in the immediate future, combined with a plan of action; (d) action, marked by specific activities and behavioral change; and finally (e) maintenance with ongoing commitment to maintain the effects of the change (Burbank, Padula & Nigg, 2000; Velicer, Prochaska, Fava, Norman & Redding, 1998). This theoretical model has proven successful in developing and evaluating a wide range of health interventions including smoking cessation, exercise, weight loss, medical compliance, and consistent use of sunscreen (Kristjánsson, Ullén & Helgason, 2004; Prochaska, DiClemente & Norcross, 1992; Takeuchi, Hillers, Edwards, Edlefsen & McCurdy, 2005; Velicer et al., 1998).

Along with investigations of fall prevention, there is a significant body of research that examines the relationship between balance self-efficacy and the fear of falling among older adults. Community-dwelling older adults may avoid certain household tasks due to a fear of falling. In one study, nearly 25% admitted avoiding everyday activities due to the fear of falling (Tinetti et al., 1990). Self-efficacy refers to an individual’s perception of his or her own capabilities inside a specific domain of activities (Bandura, Adams, & Beyer, 1977; Bandura, 1994). Thus, in addition to measuring actual change and intentions to change in the area of fall prevention, self-efficacy for maintaining one’s balance could prove to be a useful secondary measure of intention to institute fall prevention strategies. We believe there is a reciprocal relationship between balance self-efficacy and efforts to reduce the risk for falling since behavioral change can be influenced through elevated self-efficacy (Bandura, Jeffrey, & Gajdos, 1975). Efforts to improve the safety of their environment should increase balance self-efficacy in older adults. Also, higher initial levels of balance self-efficacy may help individuals take the steps needed to prevent future falls, due to higher levels of confidence in their capability to engage with key aspects of their environment.

This study examined the effects of a fall prevention program consisting of mailed information and suggestions for bathroom safety modifications. Our hypotheses were:

1. Compared to those in the control group, participants who received the mailed intervention materials would be more likely to report both an intention to change, and actual installation of specific bathroom safety devices (i.e., grab bars, bathtub mats, and taped bathroom rugs).

2. Across all study participants, balance self-efficacy would be higher among individuals stating an intention to change (contemplation and preparation stages of change) than among those in the precontemplation stage of change.

Method

Participants

We selected individuals aged 65–97 years ($M = 77.7$ years, $SD = 6.6$) from a larger pool of 303 community-dwelling older adults who had completed a previous telephone survey on home modifications. That original sample had been randomly drawn from registered voters living in a specific census district with a high

physician (Chou, Tinetti, King, Irwin, & Fortinsky, 2005).
 proportion (21%) of older adults, and who had agreed to further contact regarding research participation. Individuals who had reported in the earlier survey that their bathroom already had grab bars installed were considered ineligible to participate in the current study. Thus, from the original pool of 303, a subset of 176 older adults was identified as eligible for this research. Data from the initial survey indicated that in this subset, 85% of the participants occupied a single-family home (owned or rented) in comparison to a national average of 80.5% home ownership for the identical age bracket of persons age 65 and older (U.S. Census Bureau, 2007).

We randomly assigned these eligible individuals (N = 176) to either the intervention group that received mailed prevention materials (n = 87), or the assessment-only control group that did not receive any prevention materials (n = 89). We did not provide monetary compensation to any participants in either group. At the time of the study, each eligible participant lived independently in the community and provided informed consent via the telephone immediately prior to the telephone interview used to assess outcome.

Measures

Installation of bathroom safety devices. In the outcome assessment conducted via telephone interview, we asked participants, “Do you currently have installed in your bathroom a grab bar device?” If they indicated a grab bar was installed, we coded them as being in the action or maintenance stage according to the length of time that the grab bar was installed: last 30 days (action) or 31 days to 12 months (maintenance). The interview continued with the same question regarding installation of a nonskid surface for the tub/shower, and bathroom floor mats that are taped to the floor. For the purposes of the present study, we treated installation (by the participants themselves or someone else under their direction) of any of the three safety devices within the previous 30 days (action stage) as the primary outcome measure.

Intent to install safety devices. Using the Transtheoretical Model of Change (TTM, Prochaska & DiClemente, 1982), we constructed four items that reflect the continuum of intentional change. We asked those participants without safety devices already installed in their bathrooms: “Are you intending to install any type of grab bar device, non-slip surface in the tub/shower, or planning to tape down the rugs in your bathroom?” We coded a “no” response as the individual being in the pre-contemplation stage; we followed a “yes” response with an open-ended question asking about the perceived time frame for future change. We coded those stating an intention to install some form of a safety device at some point in the future to be in the contemplation stage, and those stating a serious intent to install a specifically-named safety device within the next 3 months to be in the preparation stage.

Activities-specific Balance Confidence Scale. This 16-item self-efficacy scale (Powell & Myers, 1995) measures self-perceived confidence to engage in specific activities “without losing your balance or becoming unsteady” (p. M30). Participants rate their confidence for each item on a continuous scale from 0% to 100% with 0% representing no confidence and 100% representing complete confidence in performing the activity without becoming unsteady. An example question is “How confident are you that you can pick up a slipper from the floor without losing your balance or becoming unsteady?” (Powell & Myers, 1995). This scale demonstrates reliability, validity, and responsiveness in assessing an elderly person’s fear of falling and has increased sensitivity over other efficacy measures (Jørstad et al., 2005).

Procedures

Intervention components. We mailed educational materials via U.S. Postal Service directly to the homes of the participants in the intervention group. The educational intervention packets included an envelope bearing a return address label of the investigator’s university.

There were four informational elements to the educational intervention. The first component of the mailed packet consisted of an introductory letter from a local community partnership, cosigned by the investigators, the local Fire Department Chief, and the local community betterment association. We obtained signed support for the project from these local municipal leaders to increase the authenticity of the study for these community dwelling seniors (Carp, 1989). The letter outlined the consequences of a severe accidental fall, statistics associated with functional decline as a result of falling, the financial expenses that could be incurred following an accidental fall, and a listing of the risk reduction behaviors known to be effective in reducing the risks for falls while promoting bathroom safety.

The second piece of the packet was a single-page, four-color Bathroom Safety Checklist developed by the investigators. It featured two separate images of older persons and a series of questions intended to spark an inventory of the participants’ risks in the bathroom and a current safety status assessment. This checklist also included a series of Action Steps for installation of the bathroom safety items. The phone number to the investigator’s lab was provided in the case of questions.

The third element of the mailed intervention was a detailed Centers for Disease Control (CDC) 14-page color brochure specifically focused on fall prevention
information and suggestions. It included specific targeted information regarding the installation of bathroom safety devices.

The final component of the educational intervention included a custom-designed store coupon provided by a locally owned and operated family hardware store. The authorized redemption of the coupon at the point of sale was limited to purchases related to bathroom safety: any type of grab bar, any type of nonskid surface device for use in the tub/shower floor, and double-sided tape for securing throw rugs in the bathroom. The promotional price reduction included an expiration date 10 weeks from the time of the mailing. The control group received none of these materials or contact of any kind from the researchers prior to the telephone interview to assess outcome.

Assessment Strategy. Following a 4-week waiting period, we conducted telephone interviews over the course of 10 calendar days to assess bathroom modifications and balance self-efficacy since a telephone interview is a viable means of data collection in this age group (Herzog & Kulka, 1989). We mailed printed questionnaires to participants unavailable by telephone. An investigator blind to the respondent’s intervention or control group designation interviewed all participants using the identical questionnaire. If a participant expressed confusion during any portion of the interview, the interviewer asked permission for a licensed clinical geropsychologist to recontact the participant and assess the need for referrals for cognitive assessment and/or intervention.

Results
Of the 176 participants in the eligible sample, 125 (71%) provided informed consent for the telephone outcomes assessment, 14 (8%) refused consent to participate in the telephone assessment, 9 (5%) had a disconnected telephone, 9 (5%) were deceased, 5 (3%) relocated, and 14 (8%) were hospitalized, out of town, or otherwise unavailable. Of those providing informed consent, 58 were in the control condition, and 67 were in the intervention condition. Rates of consent and completion of the outcomes assessment did not differ between the two conditions, \( \chi^2 (1, N = 176) = 1.59, p = .21 \).

Actual Change Defined by Completed Installation of Any Safety Device
The proportion of intervention participants who completed a safety device installation during the 30-day waiting period was larger \( (p = .08) \) than that of the control group \( (p = .01) \). A chi-square test determined whether individuals who received the mailed intervention materials were statistically more likely than control participants to have installed a grab bar or another bathroom safety device in the previous month (i.e., since receiving the mailing). The intervention provided a small, yet statistically significant behavioral change during the intervention period that consisted of installing a grab bar or another safety device in the bathroom, \( \chi^2 (1, N = 125) = 4.0, p \leq .05 \).

Intent to Change
The next set of analyses used the responses of those participants who had not yet made any changes to improve bathroom safety \( (n = 115) \). Intervention participants displayed a statistically significant increase \( (p = .39) \) over the control group \( (p = .14) \) for stating an intent to take action to reduce risks associated with falling in the bathroom, consistent with reaching a minimum of the contemplation stage of change, \( \chi^2 (1, N = 115) = 8.9, p < .01 \). Intervention participants also displayed a statistically significant increase \( (p = .12) \) over the control group \( (p = .0) \) for serious intent to take action within the next three months, consistent with the preparation stage of change, \( \chi^2 (1, N = 115) = 7.1, p \leq .01 \).

Association Between Intent and Self-Efficacy for Balance
Pooling together the intervention and control participants, those who stated an intention to engage in further installation of any of the three safety devices, consistent with being in either the contemplation or preparation stages of change, reported higher self-efficacy, \( t (1, 92) = 1.9, p \leq .05 \), than those who were coded as being in the precontemplation stage of change (i.e., no intention to make any further safety installations). Because the self-efficacy scores of those in the precontemplation stage varied more than those of the other participants, equal variances were not assumed in the statistical analyses. There was a higher level of balance self-efficacy for those intending to install a bathroom safety device \( (n = 33, M = 80.0, SD = 12.6) \), compared to those not intending to install a bathroom safety device \( (n = 92, M = 74.0, SD = 20.5) \).

Discussion
The results support our Hypothesis 1 regarding the positive impact of the intervention on both intent and actual installation of bathroom safety features. Compared to control group participants, older adults receiving educational materials about bathroom safety had significantly higher levels of both intent and serious intent to improve the safety of their own bathrooms. Intervention group participants were more likely than control participants to be in either the contemplation or preparation stages of change; the control group included the largest proportion precontemplators.
Intervention participants also had higher numbers of completed bathroom safety modifications. In addition to demonstrating an impact for the intervention, this study lends further support to the Transtheoretical Model of Change (TTM, Prochaska & DiClemente, 1982) which views levels of intention and action as a process measured along a continuum.

We also found support for Hypothesis 2; regardless of intervention group or control group status, those participants who had higher self-efficacy for maintaining their balance in a range of situations were more likely to state their intent to make bathroom safety improvements. However, we do not have the ability to disentangle the direction of these effects; declaring a plan to make safety improvements may lead to an increase in balance self-efficacy, or higher pre-existing levels of balance self-efficacy may help individuals to consider environmental modifications. Efficacy expectations rise and fall with self-regulated motivation and the expectation that a particular activity will likely result in a successful outcome (Bandura, 1994). Increased adherence to behavioral change has been associated with additional home modifications and maintenance of fall risk reduction behaviors (Finkelstein et al., 2004; Li et al., 2001). Yardley and Kempen (2006) state that clinical interventions designed to reduce the risk for falls can improve self-efficacy via strategies that include environmental improvements. And, it is also the case that persons with a low perceived self-efficacy have an increased tendency to avoid specific activities (e.g., reach at eye level, walk about the house, get in/out of a car, sweep the floor) considered essential for everyday living (Powell & Myers, 1995). Future research that specifically enrolls participants on the basis of low levels of balance self-efficacy may help demonstrate that the intervention has a direct impact on increasing levels of self-efficacy, as well as influencing fall safety behaviors.

Because individuals with higher self-efficacy appear more likely to plan environmental renovations, as suggested by our findings, how can community dwelling older adults with lower self-efficacy be encouraged to make the necessary changes to reduce the risks associated with falling in the bathroom? Bandura’s (1994) self-efficacy theory suggests that perceived similarities between the individual and role models allow for greater influence on both beliefs and behaviors. Future interventions might use existing social networks of older friends and family members who are already engaging in safety behaviors as role models for home improvements.

There were some measurement issues that could be improved in future studies of balance self-efficacy of older adults. Lach (2006) suggests that using Bandura’s theory and measuring outcome expectancies with the Activities-specific Balance Confidence Scale (ABC, Powell & Myers, 1995) can offer an expanded understanding of the reciprocal associations between self-efficacy beliefs and safety behaviors. Adding bathroom-specific items on the balance self-efficacy instrument would improve outcome assessment. A sample question might be, “How confident are you that you can step out of the bathtub or shower without holding onto the wall or towel rod without losing your balance or becoming unsteady?” This level of specificity in measurement would be helpful in research that targets bathroom safety and fall prevention.

Because this study focused primarily on risk reduction associated with falling in the bathroom, we cannot speak to the direct costs and benefits associated with the change experienced by the participants. However, the CDC asserts that the average cost per fall injury in 2000 was $17,500, excluding doctors’ fees (Roudsari, Ebel, Corso, Molinari, & Koepsell, 2005), and predicts that by 2020 the national costs associated with adult falls to reach $54.9 billion (Englander, Hodson, & Terregrossa, 1996). Even small changes associated with preventive interventions have the capacity to produce a significant impact on health care costs.

Further study is warranted to determine what is necessary for transition from intent to change to increased actual behavioral change, defined by completed installation of at least one bathroom safety device. It might be the case that combining mailed educational materials with a local vocational education center or trade union apprenticeship program that offered free, supervised installation of bathroom safety features would increase levels of behavioral change. Targeting geographic regions having the largest number of older residents with strategically placed billboards, public service announcements, or advertising on or near public transportation would move this intervention from the individual to the community public health level.

References
Improving Bathroom Safety | Murphy and Steffen

Journal of General Internal Medicine, 21, 117-122.
Being a parent is stressful. Many demands are placed on parents today at home, at work, and at children’s extra-curricular activities. Family functioning is greatly influenced by parents’ ability to meet the many demands placed upon them both inside and outside of their parenting roles (McCubbin and Patterson, 1983). Parenting stress may come directly from issues with the child, such as child characteristics (e.g. distractibility, temperament, demandingness; Abidin, 1983), and physical care requirements; or from the parenting role itself, such as from restrictions caused by the parenting role, the ability of the parents to take care of their own needs, and strains in the marital relationship. Parenting stress issues change with the maturing child. For example, childhood issues focus on direct care demands, whereas adolescence requires accepting increased demands for independence. With adolescents parents must cope with the adolescents’ ability to independently make wise life decisions in terms of goals, the people they choose to relate with, how they choose to spend their free time away from them unsupervised, and the ways they interact in the world. Allowing and accepting this independence may cause significant stress for parents of adolescents (Sheras, Abidin, & Konold, 1998).

McCubbin and Patterson’s (1983) Double ABCX Theory of Family Stress suggests that it isn’t necessarily a single event that causes stress for parents, as much as many demands that pile up and overextend the resources available to the family. An event or pileup of events (Factor A), creates stress for the family to the extent that it exceeds the family’s available resources (Factor B) to overcome the stressor and prevent a crisis from occurring. The perception of the ability to use available resources (Factor C), along with the family meaning assigned to the stressor, and behavioral responses to the stressor all combine to determine the outcome (Factor X) of the stressor upon the family. Demand-resource imbalance can result in dysfunctional family responses to stress depending on the nature of

Changes in Parents’ Stress as Their Children Become Adolescents: A Validation of the Stress Index for Parents of Adolescents

This longitudinal assessment of changes in mothers’ and fathers’ parenting stress from their children’s mid-childhood to adolescence provides a unique inspection of the changes in stress experienced by both parents as children mature. Fifty-six mothers and fathers reported on parenting stress levels when their children were eight years of age and again when these children were 14 years of age. The Stress Index for Parents of Adolescents (SIPA, Sheras, Abidin, & Konold, 1998) was found to be a valid upward extension of the Parenting Stress Index (PSI, Abidin, 1983). Results of the current study suggest relative stability of parenting stress as children mature from childhood to adolescence and show few differences in the stress levels experienced by mothers and fathers.

Melissa K. Wheatley
Diane E. Wille
Indiana University Southeast

Author Note. Melissa K Wheatley, School of Social Sciences, Indiana University Southeast, 4201 Grant Line Road, New Albany, IN 47150. Email: melissa.wheatley1@gmail.com
Reprint requests should be sent to Diane E. Wille, School of Social Sciences, Indiana University Southeast, 4201 Grant Line Road, New Albany, IN 47150. Email: dwille@ius.edu.

Acknowledgements. Preparation of this article was made possible by Indiana University Southeast Fellowship Awards to the first author. Special thanks to the families who participated in this research project and to the following research assistants: Shana Weddle, Bethany Craig, and Christy Guifo.
the situation, family characteristics, and psychological and physical well-being of family members.

Few studies on parenting stress are longitudinal and this is the first study, to date, to evaluate parenting stress from childhood to adolescence using the Parenting Stress Index (PSI, Abidin, 1983) and the Stress Index for Parents of Adolescents (SIPA, Sheras, Abidin, & Konold, 1998). The PSI and the SIPA are assessments used by clinicians and researchers to measure the level of distress experienced by parents. The PSI has proven to be a valid assessment of the parenting stress of parents with children up to 12 years of age (Abidin, 1983); the recently developed SIPA is designed for assessing parenting stress of parents with children 12 years and older (Sheras et al., 1998). The SIPA was created as an upward extension of the PSI and has limited validation research. An advantage in using both the PSI and SIPA to measure longitudinal parenting stress is that these measures provide continuity by assessing similar constructs while also taking into account changing issues of adolescence that other measures do not assess.

Parenting Stress During Childhood and Adolescence

Longitudinal studies of parenting stress across childhood suggest relative stability of parenting stress (Beckman, 1991; Belsky & Hsieh, 1998; Deater-Deckard, 2004; Dyson, 1993). For example, moderate correlations have been found in parenting stress reported when children were 18 months old and again when these children were 72 months old (Beckman, 1991), and for parents with children entering preschool and again when their children entered grade school (Belsky & Hsieh, 1998). This stability has also been found in parents of children with disabilities (Dyson, 1993).

In accordance with the Double ABCX Family Stress Model, parents’ sense of their ability to competently handle their life responsibilities (their available resources) may have an influence on the parenting stress experienced (McCubbin & Patterson, 1983). In today’s society with more mothers working and fathers taking more responsibility for the care of the children, the amount of the household workload and childcare responsibility shared between members of the family can greatly affect family functioning (Belsky & Hsieh, 1998; Dempsey, 2000). Increased child care demands can increase parenting stress for mothers and decrease family harmony (Keller & Honig, 2004). If parents are not content with the balance of roles and their ability to successfully fulfill their family duties, stress can result (McCubbin & Patterson, 1983). Mothers receiving more emotional and physical help from their spouse in caring for the child report less parenting stress directly related to parental responsibilities (Willinger, Diendorfer-Radner, Willnauer, Jögl, & Hager, 2005).

Research suggests that equality of household duties doesn’t affect maternal parenting stress levels as much as division of childcare (Belsky & Hsieh, 1998).

Is parenting stress stable as the child matures or does parenting stress change along with changing parenting issues as children become adolescents? A lack of empirical research on parenting stress experienced by parents of adolescents impedes our knowledge about the process by which parenting stress changes as the child matures into adolescence. A review of the literature on changes in the experiences of parents at different stages of parenthood suggests that parental competency becomes an issue during adolescence due to parents’ increased loss of control of their adolescent and allowing their adolescent to become more independent (Pasley & Gecas, 1984; Ballenski-Bracket & Cook, 1982). One would expect a reduction in physical childcare needs during adolescence, but at the same time complex challenges arise in maintaining the proper balance between protecting and encouraging the independence of the adolescent.

Comparison of Parenting Stress of Mothers and Fathers During Childhood and Adolescence

When comparing the parenting stress of mothers and fathers as assessed by the PSI, few overall parenting stress differences have been found (Beckman, 1991; Dyson, 1993; Hadadian, 1994; Keller & Honig, 2004; Putnick et al., 2008), but differences have been found in the domains and subscales. Fathers have been found to experience more stress related to child characteristics such as child temperament, lack of feeling rewarded by interactions with the child (Krauss, 1993; Roach, Orsmond, & Barratt, 1999), and stress related to the attachment relationship with the child (Keller & Honig, 2004; Krauss, 1993). Mothers have been found to experience more stress related to the ability to take care of their own health needs, restrictions imposed by the parenting role, and the spousal relationship (Beckman, 1991; Krauss, 1993).

We found no published longitudinal studies that assessed gender differences in parenting stress using the SIPA. In a cross-sectional study of parenting stress of children between 10 and 17 years of age, Small, Eastman, and Cornelius (1988), found no differences in parental stress between mothers and fathers, but they found that fathers had higher parental stress due to adolescent non-adherence to parental advice and adolescent deviant behaviors. Mothers’ parental stress was higher when adolescents desired more autonomy. This is consistent with childhood PSI results that suggest higher stress in fathers due to child characteristics (Beckman, 1991; Keller & Honig, 2004; Krauss, 1993; Roach, Orsmond, & Barratt, 1999) and in mothers.
due to their parental role, although the source of this parental role stress appears to change from issues of direct care in childhood, to accepting more autonomy in their adolescent children.

**Purpose of Current Study**

To date, few studies have been published which have used the SIPA and very little validity information exists. A search of the psychology article databases revealed only 11 articles using the SIPA. Based on previous research, continuity is expected in the parenting stress of parents with children between 8 and 14 years of age. The current findings will help provide evidence of the validity of the SIPA as an upward extension of the PSI and will provide a rare 6-year longitudinal analysis of changes in parenting stress across this transitional time. The experimenters also expected that differences in mothers and fathers responses on the SIPA would reflect those differences previously found on the PSI. Overall level of parenting stress will be similar for mothers and fathers; however, mothers will report greater parental role stress and fathers greater stress due to adolescent characteristics and relationship with adolescents. Consistent with the Double ABCX Theory (McCubbin & Patterson, 1983) it is also expected that parents at both time periods will report more stress when the parents report less equitable childcare.

**Method**

**Participants**

Seventy Midwestern families were recruited to participate in a longitudinal study of family relationships through public birth announcements in local newspapers. Letters sent to the families were followed up with a phone call inviting their participation. The families were first visited when the target child was 6 months of age. Fourteen families dropped out of the study for various reasons over the 14-year study duration. The current study is based on 56 families who participated in both the mid-childhood and adolescent sessions. There were no significant differences in the information collected during the early home visits between the families who dropped out and those who remained in the study. Eighty-two percent of the families were intact at each session. In two families parents divorced before the mid-childhood session and remarried by the adolescent session and one father was deceased by the adolescent session.

The average age of the target children (26 girls and 30 boys) was 7.76 years (SD = .36 years) at the mid-childhood visit and 14.49 years (SD = .36 years) at the adolescent visit. The average age of the fathers was 38 years (M = 38.18 years, SD = 3.66 years) at the mid-childhood visit and 45 years (M = 44.91 years, SD = 3.68 years) at the adolescent visit. The total number of children in the families, including the target child, ranged from 1 to 5 (Mode = 2). Thirty-two percent of the target children were first-born, 9% were only children, and 59% were later born. All participants were European American. Socioeconomic status ranged from low-middle to middle class (M = 45.94 SD = 9.65) on the Hollingshead 4-factor Index (Hollingshead, 1978).

**Materials**

**Parenting Stress Index.** The PSI consists of 120 questions measuring parenting stress in three domains (parent, child, and life stress). Parents respond using 5-point Likert scales ranging from *strongly disagree* to *strongly agree*. The PSI domain and subscale scores were computed using the standard procedure developed by Abidin (1983). The parent and child domain scores were combined to provide a total score for parenting stress. High scores on the child domain suggest high stress due to child characteristics, high scores on the parent domain suggest high stress due to parental functioning, and high scores on the optional Life Stress Domain suggest high stress due to factors outside of the parenting role. The internal consistency of the domains and subscales of the PSI were assessed using coefficient alphas.

Over the last two and a half decades studies have provided evidence for the validity of the PSI (Abidin, 1983) for use by parents of both genders, in a wide variety of parental circumstances and demographic contexts. For example, high levels of stress related to characteristics of the child have been found among parents of children with craniofacial birth defects (Speltz, Armsden, & Clarren, 1990) and mothers of boys diagnosed with pervasive hyperactivity (Beck, Young, & Tarnowski, 1990). Convergent validity for the PSI is supported by a comparison of parents’ responses on the PSI and Questionnaire on Resources and Stress (QRS; Friedrich, Greenberg, & Crnic, 1983; Sexton, Burrell, Thompson, & Sharpton, 1992). The correlation between the total stress score from the PSI and the QRS was .65 (p < .001) (Sexton et al., 1992).

**The Stress Index for Parents of Adolescents.** The SIPA (Sheras, Abidin, & Konold, 1998) was developed as an extension of the PSI to reflect the changing developmental issues faced by adolescents and their parents (Sheras et al., 1998) and it is used for parents of both genders in a wide variety of parental circumstances and demographic contexts. The SIPA consists of 112 items measuring 4 domains that were constructed using
methods similar to those used to create the PSI. Many of the items in the Parent Domain were retained from the PSI, because these items are not affected by developmental maturity of the child. Factor analysis resulted in 4 subscales in the Adolescent Domain, 4 subscales in the Parent Domain, and a Life Stress Domain with no subscales (Sheras et al., 1998). Higher order factor analysis also revealed an Adolescent-Parent Relationship Domain.

Parents respond to items on the SIPA using 5-point Likert scales ranging from strongly disagree to strongly agree. The SIPA domain and subscales scores were computed using the standardized procedure developed by Sheras et al. (1998). The parent, adolescent, and adolescent-parent relationship domain scores were combined to give a total score for parenting stress. High scores on the adolescent domain suggest high stress caused by adolescent characteristics; high scores on the parent domain suggest high stress caused by parental functioning; high scores on the adolescent-parent relationship domain suggest high stress caused by the relationship between parent and adolescent; and high scores on the optional life stress domain suggest high stress due to factors outside of the parenting role.

Procedure
During the mid-childhood home visits parents completed the PSI (Abidin, 1983), and the SIPA (Sheras et al., 1998) was completed during the adolescent visit. During both home visits parents responded to questions about age, number of children in the family, education, occupation, and percentage of child care responsibility provided by both mother and father.

Results
Continuity of Parenting Stress Between Mid-Childhood and Adolescence
The continuity of parenting stress when their children move from mid-childhood to adolescence was assessed using Pearson correlations for the major domains (child-adolescent and parent-parent) and Total Stress only. Due to differences in the number of items in the scales of the PSI and the SIPA, the means for the major domains and total stress were computed and used in this analysis. For both mothers and fathers, there were significant positive correlations between the mid-childhood and adolescence in Child/Adolescent Domains \( r(52) = .46, p < .05; r(45) = .56, p < .01 \) for mothers and fathers respectively, the Parent Domains \( r(51) = .59, p < .01, r(46) = .52, p < .01 \) for mothers and fathers respectively, and Total Stress \( r(51) = .40, p < .05, r(45) = .58, p < .01 \) for mothers and fathers respectively. These moderately high correlations provide evidence of the continuity of parenting stress between childhood

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mothers</th>
<th>Fathers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child Domain</td>
<td>.77</td>
<td>.81</td>
</tr>
<tr>
<td>Adaptability</td>
<td>.49</td>
<td>.60</td>
</tr>
<tr>
<td>Acceptability</td>
<td>.52</td>
<td>.70</td>
</tr>
<tr>
<td>Demandingness</td>
<td>.29</td>
<td>.40</td>
</tr>
<tr>
<td>Mood</td>
<td>.56</td>
<td>.68</td>
</tr>
<tr>
<td>Distractibility/Hyperactivity</td>
<td>.67</td>
<td>.39</td>
</tr>
<tr>
<td>Reinforces Parent</td>
<td>.52</td>
<td>.67</td>
</tr>
<tr>
<td>Parent Domain</td>
<td>.84</td>
<td>.89</td>
</tr>
<tr>
<td>Depression</td>
<td>.66</td>
<td>.70</td>
</tr>
<tr>
<td>Attachment</td>
<td>.47</td>
<td>.60</td>
</tr>
<tr>
<td>Role Restriction</td>
<td>.67</td>
<td>.75</td>
</tr>
<tr>
<td>Competence</td>
<td>.65</td>
<td>.68</td>
</tr>
<tr>
<td>Isolation</td>
<td>.70</td>
<td>.71</td>
</tr>
<tr>
<td>Spouse</td>
<td>.59</td>
<td>.65</td>
</tr>
<tr>
<td>Health</td>
<td>.62</td>
<td>.57</td>
</tr>
<tr>
<td>Total Stress</td>
<td>.86</td>
<td>.91</td>
</tr>
</tbody>
</table>

*Note: numbers are coefficient alphas.*

**TABLE 2**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mothers</th>
<th>Fathers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adolescent Domain</td>
<td>.94</td>
<td>.95</td>
</tr>
<tr>
<td>Delinquency</td>
<td>.86</td>
<td>.85</td>
</tr>
<tr>
<td>Moodiness/Emotional Lability</td>
<td>.90</td>
<td>.91</td>
</tr>
<tr>
<td>Failure to Achieve</td>
<td>.92</td>
<td>.91</td>
</tr>
<tr>
<td>Isolation</td>
<td>.88</td>
<td>.87</td>
</tr>
<tr>
<td>Parent Domain</td>
<td>.91</td>
<td>.88</td>
</tr>
<tr>
<td>Life Restrictions</td>
<td>.88</td>
<td>.87</td>
</tr>
<tr>
<td>Relationship with Spouse</td>
<td>.87</td>
<td>.80</td>
</tr>
<tr>
<td>Social Isolation</td>
<td>.77</td>
<td>.72</td>
</tr>
<tr>
<td>Incompetence</td>
<td>.73</td>
<td>.78</td>
</tr>
<tr>
<td>Adolescent-Parent Relationship</td>
<td>.81</td>
<td>.81</td>
</tr>
<tr>
<td>Total Stress</td>
<td>.94</td>
<td>.95</td>
</tr>
</tbody>
</table>

*Note: numbers are coefficient alphas.*
and adolescence and provide support for the validity of the SIPA as an upward extension of the PSI.

The PSI subscale internal consistencies were lower than reported by Abidin (1983) but comparable to recently published subscale internal consistencies (Halme, Tarkka, Nummi, & Astedt-Kurki, 2006); however, the internal consistencies for the domains and total stress were comparable (Table 1). Internal consistency of the domains and subscales of the SIPA was assessed using coefficients alphas (Table 2) and are comparable to the internal consistencies reported by Sheras et al. (1998).

Comparison of Mothers’ and Fathers’ Parenting Stress

Two by two Analyses of Variance (ANOVA) with gender of parent and gender of child as the independent variables were performed on all PSI and SIPA domains and subscales (Tables 3 and 4 respectively). At 8 years, Attachment stress and Relationship with Spouse stress were significantly different for mothers and fathers. Fathers reported greater stress in their Attachment relationship with their 8-year-olds than mothers, whereas mothers reported greater stress in Relationship with Spouse than fathers. At 14 years, Incompetence differentiated mothers and fathers. Mothers reported greater Incompetence than fathers.

Few gender of child effects were found. Parents of 8-year-old sons reported higher stress on Distractibility/Hyperactivity, \( F(1, 55) = 5.28, \ p < .05 \), than parents of daughters. When children were 14 years old, parents of sons reported more Isolation, \( F(1, 46) = 3.10, \ p < .05 \), and Achievement stress, \( F(1, 46) = 6.45, \ p < .05 \), than parents of daughters.

Equitable Childcare Responsibility and Parenting Stress

To assess the influence of equitable childcare responsibility on parenting stress, families were grouped based upon the percentage of childcare responsibility reported by the mothers and fathers. During the interview the parents were asked about the percent of child care responsibility for each of them. The parents agreed on a percentage. Typically this agreement was immediate and no disagreements were noted. The High Equitability Group contained families with equitable childcare responsibility in which parents reported the father provided 50% or more of the childcare responsibility. The Mildly Discrepant Group consisted of families in which the father provided between 31-49%

**TABLE 3**

<table>
<thead>
<tr>
<th>Domain/Subscale</th>
<th>( F )</th>
<th>( p )</th>
<th>( M )</th>
<th>( SD )</th>
<th>( M )</th>
<th>( SD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child Domain</td>
<td>0.00</td>
<td>.96</td>
<td>1.99</td>
<td>.04</td>
<td>2.00</td>
<td>.03</td>
</tr>
<tr>
<td>Distractibility</td>
<td>0.13</td>
<td>.72</td>
<td>2.56</td>
<td>.07</td>
<td>2.43</td>
<td>.07</td>
</tr>
<tr>
<td>Adaptability</td>
<td>0.12</td>
<td>.73</td>
<td>2.15</td>
<td>.06</td>
<td>2.17</td>
<td>.05</td>
</tr>
<tr>
<td>Reinforces Parent</td>
<td>1.65</td>
<td>.20</td>
<td>1.47</td>
<td>.05</td>
<td>1.57</td>
<td>.06</td>
</tr>
<tr>
<td>Demandiness</td>
<td>1.56</td>
<td>.22</td>
<td>1.92</td>
<td>.05</td>
<td>1.84</td>
<td>.04</td>
</tr>
<tr>
<td>Mood</td>
<td>0.08</td>
<td>.79</td>
<td>1.89</td>
<td>.07</td>
<td>1.92</td>
<td>.07</td>
</tr>
<tr>
<td>Acceptability</td>
<td>0.00</td>
<td>.98</td>
<td>1.64</td>
<td>.05</td>
<td>1.64</td>
<td>.06</td>
</tr>
<tr>
<td>Parent Domain</td>
<td>0.83</td>
<td>.37</td>
<td>2.07</td>
<td>.04</td>
<td>2.02</td>
<td>.04</td>
</tr>
<tr>
<td>Competence</td>
<td>0.38</td>
<td>.54</td>
<td>1.96</td>
<td>.05</td>
<td>1.91</td>
<td>.05</td>
</tr>
<tr>
<td>Isolation</td>
<td>2.22</td>
<td>.14</td>
<td>1.96</td>
<td>.06</td>
<td>2.07</td>
<td>.07</td>
</tr>
<tr>
<td>Attachment</td>
<td>4.27</td>
<td>.04</td>
<td>1.61</td>
<td>.05</td>
<td>1.74</td>
<td>.05</td>
</tr>
<tr>
<td>Health</td>
<td>2.04</td>
<td>.16</td>
<td>2.19</td>
<td>.08</td>
<td>2.20</td>
<td>.08</td>
</tr>
<tr>
<td>Role Restrictions</td>
<td>0.58</td>
<td>.45</td>
<td>2.30</td>
<td>.07</td>
<td>2.23</td>
<td>.08</td>
</tr>
<tr>
<td>Depression</td>
<td>0.99</td>
<td>.32</td>
<td>2.08</td>
<td>.06</td>
<td>2.00</td>
<td>.06</td>
</tr>
<tr>
<td>Relationship with Spouse</td>
<td>8.83</td>
<td>.00</td>
<td>2.43</td>
<td>.07</td>
<td>2.19</td>
<td>.07</td>
</tr>
<tr>
<td>Total Stress</td>
<td>0.31</td>
<td>.58</td>
<td>2.03</td>
<td>.03</td>
<td>2.01</td>
<td>.04</td>
</tr>
</tbody>
</table>
of the childcare responsibility. The Low Equitability Group is defined as those with fathers providing 30% or less of the childcare responsibility.

Two-way ANOVAs with gender of parent and childcare responsibility group as the independent variables and PSI and SIPA domains and subscales as the dependent variables revealed that equitable childcare involvement had different effects on maternal and paternal stress. At 8 years, equitable childcare involvement reduced maternal stress. Mothers in Highly Equitability families reported less Total Stress, $F(2, 59) = 3.24, p < .05$, Parent Domain stress, $F(2, 61) = 4.13, p < .05$, and less Parent Domain subscale stress on Competence, $F(2, 61) = 3.81, p < .05$, Isolation, $F(2, 59) = 28.20, p < .05$, and Relationship with Spouse, $F(2, 59) = 7.62, p < .01$. At 14, equity of childcare had no effect on maternal stress. Equitable childcare had no effect on paternal stress at 8 years. At 14 years paternal involvement reduced paternal stress. Fathers in High Equitability families reported less Incompetence stress, $F(2, 47) = 3.34, p < .05$.

**Discussion**

In agreement with previous longitudinal research on parents of children (Beckman, 1991; Dyson, 1993), the results of the current study suggest a strong continuity of parenting stress as children move from childhood to adolescence. The moderately large correlations in parenting stress between 8 and 14 years on the PSI and SIPA provide evidence of the validity of the SIPA as a true upward extension of the PSI. Family practitioners should feel confident in the validity of the SIPA in assessing parenting stress when the PSI is no longer useful due to the changing issues that arise with child maturation.

The few gender of parent differences that were found, i.e., maternal relationship with spouse stress and paternal attachment stress when child is 8 years old, and maternal incompetence when child is 14 years old, are consistent with previous research (Beckman, 1991; Hadadian, 1994; Keller & Honig, 2004; Krauss, 1993). These differences provide insight into family dynamics and the relative relationships of family members. When children are young, mothers are stressed over their relationships with their spouses while their spouses are stressed over their relationships with the children. As the child matures, mothers are then stressed over their competence as a parent as the adolescent starts to break away from the family.

Parental and societal resistance to role changes within the family may be a problem facing families today. Despite the increased employment of mothers, there remains the persistent view that women are more appropriate primary childcare providers (Hill, Hawkins, Alan, Martinson, & Ferris, 2003). The current findings suggest that maternal stress is lower and there is no difference in paternal stress when there is more equitable childcare responsibility by mid-childhood. Deater-Deckard and Scarr (1996) warn that equal division of parental labor only eases parenting stress to

<table>
<thead>
<tr>
<th>Stress Index for Parents of Adolescents and Gender of Parent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain/Subscale</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Adolescent Domain</td>
</tr>
<tr>
<td>Mood</td>
</tr>
<tr>
<td>Isolation</td>
</tr>
<tr>
<td>Delinquency</td>
</tr>
<tr>
<td>Achievement</td>
</tr>
<tr>
<td>Parent Domain</td>
</tr>
<tr>
<td>Role Restrictions</td>
</tr>
<tr>
<td>Relationship with Spouse</td>
</tr>
<tr>
<td>Social Isolation</td>
</tr>
<tr>
<td>Incompetence</td>
</tr>
<tr>
<td>Adolescent-Parent Relationship</td>
</tr>
<tr>
<td>Total Stress</td>
</tr>
</tbody>
</table>
the extent that it is consistent with parental attitudes about who should be doing what. Mothers, at times, have difficulty stepping back and allowing fathers to perform childcare activities because fathers tend to do things differently than mothers. Mothers’ criticism of fathers’ parenting practices can have detrimental effects on their willingness to provide help with the child (Deater-Deckard & Scarr, 1996). For example, McBride and Rane (1998), found fathers to be less involved with, be less responsible for, interact less with, and be less accessible to the child when their spouse was less confident in their abilities.

Fathers who have more childcare involvement report better attachment to their children, the children report more security with the fathers, both parents report more happiness, and the mothers report less stress in their relationship with their spouses (Hill et al., 2003). Grych and Clark (1999) found that greater childcare involvement of fathers of infants reduced their stress, improved their relationships with their spouses, and increased family harmony. While the current findings suggest no increase or reduction of parenting stress or competence for fathers with equitable childcare responsibility of children in mid-childhood, they do indicate that equitable childcare responsibility is related to reduced stress and increased competence with adolescent children. This lack in an increase in competence with more equitable childcare in mid-childhood may depend on the mothers’ ability to step back and allow the fathers to perform childcare activities even if they do things differently.

In adolescence, when mothers are likely to be less certain about their parenting decisions, indicated by their increased competency stress, fathers may have more liberties in making decisions that result in a reduction of their parenting stress and an increase in their feelings of competency as a parent. Equitable childcare had no effect on maternal stress during adolescence, perhaps because she remains uncertain about the parental decisions being made about increased independence of the adolescent.

Implications for Practitioners
The results of the current study suggest that maternal competency stress increases as children mature from childhood to adolescence. Encouraging fathers to take more responsibility for childcare even in adolescence can help to reduce fathers’ parenting stress and increase their parental competency. Parents should be encouraged to attend parenting classes that will provide information about the normal behaviors of children at various stages of development to alleviate unnecessary fear and uncertainty about whether a given behavior is something normal or a sign of something more serious. This education could increase both parents’ competence and should also increase the mother’s perception that fathers are competent childcare providers. The competency issue for mothers of adolescents is important to address because mothers who don’t feel competent often suffer from impatience and intolerance.

Marital discord due to parenting stress and disagreements about child-care responsibility should receive particular attention by practitioners. Mothers who are experiencing depression and stress due to marital problems are more likely to perceive behavior problems in their children (Webster-Stratton, 1988) and be more punitive (Deater-Deckard, & Scarr, 1996). Stressed parents are also less responsive and have more stress in their parent-adolescent relationships which in turn has a direct negative effect on adolescent outcomes (Rothbaum & Weisz, 1994; Seginer, Vermulst, & Gerris, 2002).

Study Limitations
The sample of the current study consisted of low-middle to middle-class European-American families. We cannot state with certainty that the findings of this study would hold up with populations of different socio-economic status, or different cultures. However, the findings in this study are in line with many previous studies consisting of different populations, and with parents of children of all ages, which suggests that these findings are typical across population variables. The current study contributes to the literature by providing a cohesive longitudinal evaluation of parenting stress using two measures designed to provide a continuous measure of parenting stress but to specifically address unique aspects of parenting stress at two different stages of child development. Future studies might target different populations to investigate the similarities and differences in parenting stress longitudinally and in relation to different socio-economic statuses, cultures, or in families including step-parents.

With more mothers joining the workforce, parents need to learn to balance roles within the family. Mothers need to learn to let go of sole responsibility for childcare. More paternal childcare involvement can decrease the stressors faced by mothers who often still try to maintain their primary caregiver status and improve the father’s relationships with both the child and his spouse. The current findings suggest that the key to maintaining family harmony may lie in assisting families with balancing roles to establish an equitable balance of responsibility for the care of the children as well as financially.
Parenting Stress

References


Society Annual Convention Research Awards | Dec 1
All Psi Chi members (undergraduate and graduate) are eligible to submit their research for the Society Annual Convention Research Awards. Cash awards of $300 for undergraduates and $500 for graduates are presented to students submitting the best research for Psi Chi sessions at the APA and APS national conventions. Up to 16 awards are given: 8 for the APA Convention and 8 for the APS Convention. Awards are presented at the conventions following the presentations. Deadlines for submissions vary according to region and sometimes from year to year; check the Psi Chi website for details.

Bandura Award | Feb 1
All psychology graduate students who are Psi Chi members and graduate student affiliates of the Association for Psychological Science (APS) are eligible to submit their research for the Bandura Graduate Research Award. The winner receives the following: (1) travel expenses to attend the APS Convention to receive the award, (2) a three-year membership in APS, including subscriptions to all APS journals, and (3) two engraved plaques, one for the winner and one for the winner’s psychology department as a permanent honor to the winner. In addition, the abstract of the winning paper, as well as a photograph and brief biography of the winner, are published in the Psi Chi Newsletter.

Allyn & Bacon Awards | May 1
The Psi Chi/Allyn & Bacon Psychology Awards, cosponsored by Allyn & Bacon Publishers, are open to all undergraduate Psi Chi members and are awarded to those who submit the best overall empirical research papers. The awards are $1,000 for first place, $650 for second place, and $350 for third place. The abstracts of the winning papers, as well as photographs and brief biographies of the top three winners, are published in the Psi Chi Newsletter.

Guilford Awards | May 1
All Psi Chi undergraduate members are eligible to submit their research for the Guilford Awards. Cash awards are $1,000 for first place, $650 for second place, and $350 for third place. The abstracts of the winning papers, as well as photographs and brief biographies of the top three winners, are published in the Psi Chi Newsletter.

Research Grants
Hunt Research Grants | Oct 1
All Psi Chi student and faculty members are eligible to apply for a Hunt Research Grant. Up to three grants of up to $3,000 each are presented annually to enable members to conduct empirical research that addresses a question directly related to Psi Chi. Unlike other Psi Chi award/grant programs, the Hunt Grants focus on research directly related to the mission of Psi Chi.

SuperLab Research Grants | Oct 1
All undergraduate and graduate Psi Chi members are eligible to apply for these research grants. The purpose of this program is to provide funds to aid one undergraduate and one graduate student in conducting computer-based research. Grant winners receive a copy of SuperLab experimental lab software and a response paid from Cedrus.

Undergraduate Psychology Research Conference Grants | Oct 1
The purpose of this program is to provide funds for local/regional undergraduate psychology research conferences. Funding is intended for conferences that will invite student research presenters from at least three schools in the area and will notify all Psi Chi chapters in the geographic area of the conference. The maximum grant for each conference is $1,000.

Graduate Research Grants | Nov 1 & Feb 1
All graduate Psi Chi members are eligible to apply for these graduate research grants. All Psi Chi convention proceedings are distributed at the conventions following the presentations.

Newman Award | Feb 1
All undergraduate Psi Chi members are eligible to submit their research for the Newman Award. The winner receives the following: (1) travel expenses to attend the APA Convention to receive the award, (2) a three-year membership in APS, including subscriptions to all APS journals, and (3) two engraved plaques, one for the winner and one for the winner’s psychology department as a permanent honor to the winner. In addition, the abstract of the winning paper, as well as a photograph and brief biography of the winner, are published in the Psi Chi Newsletter.

Allyn & Bacon Awards | May 1
The Psi Chi/Allyn & Bacon Psychology Awards, cosponsored by Allyn & Bacon Publishers, are open to all undergraduate Psi Chi members and are awarded to those who submit the best overall empirical research papers. The awards are $1,000 for first place, $650 for second place, and $350 for third place. The abstracts of the winning papers, as well as photographs and brief biographies of the top three winners, are published in the Psi Chi Newsletter.

Guilford Awards | May 1
All Psi Chi undergraduate members are eligible to submit their research for the Guilford Awards. Cash awards are $1,000 for first place, $650 for second place, and $350 for third place. The abstracts of the winning papers, as well as photographs and brief biographies of the top three winners, are published in the Psi Chi Newsletter.

Regional Faculty Advisor Awards | Dec 1
All current faculty advisors and co-advisors who have served an active Psi Chi chapter for at least one year are eligible to apply for these faculty advisor research grants. The purpose of this program is to provide funds for advisors to defray the direct costs of conducting a research project. Award winners receive a check for $3,500, travel expenses to attend the APA/Psi Chi Society Annual Convention to receive the award and (2) an engraved plaque. The award is intended to recognize Psi Chi faculty advisors for their outstanding service to the chapter and to Psi Chi.

Regional Chapter Awards | Dec 1
The Psi Chi Regional Chapter Awards provide annual recognition for up to two chapters in each region that best achieve Psi Chi’s purpose. Each winning chapter receives a check for $500 and a plaque to display in the winning chapter’s department. The awards are intended to perpetuate the chapters, to identify chapters as role models for others, and to promote the purposes of Psi Chi.

Regional Faculty Advisor Awards | Dec 1
This award is presented annually to one Psi Chi faculty advisor from each region who best achieves Psi Chi’s purpose. The award is intended to recognize and rewardPsi Chi advisors who are actively involved in chapter programs. The winning faculty advisor from each region will receive $500 and a plaque.

Cousins Award | Feb 1
The Psi Chi/Ruth Hubbard Cousins Chapter Award is presented annually to the chapter that best achieves Psi Chi’s purpose. The winning chapter receives: (1) a check for $3,500, (2) travel expenses for one chapter officer to attend the APA/Psi Chi Society Annual Convention to receive the award, and (3) a plaque to display in the chapter’s department.

Website Awards | Feb 1
These awards are presented annually to three chapters with websites that are innovative, aesthetically pleasing, and that advance or support Psi Chi’s purpose. Winning chapters will receive awards of $200 each.

Kay Wilson Leadership Award | April 1
The Psi Chi/Kay Wilson Leadership Award for Outstanding Chapter Presidents is presented annually to the one chapter president who demonstrates excellence in leadership of the local chapter. The winning Psi Chi chapter officer receives: (1) a $500 cash award, (2) travel expenses for one chapter president to attend and make a short presentation at the APA/Psi Chi Society Annual Convention to receive the award, and (3) an engraved plaque commemorating the award.

Model Chapter Awards | June 30
Model Chapter Awards of $1,000 each are presented annually to recognize and reward Psi Chi chapters that consistently maintain outstanding records of membership inductions, chapter correspondence, service projects, and other criteria associated with being an outstanding chapter. All chapters submitting evidence of meeting these criteria are designated as winners.

Chapter and Advisor Awards
Denmark Award | Dec 1
The Psi Chi/Florrence L. Denmark Faculty Advisor Award is presented annually to the one Psi Chi faculty advisor who best achieves Psi Chi’s purpose. The award includes (1) travel expenses to attend the APA/Psi Chi Society Annual Convention to receive the award and (2) an engraved plaque. The award is intended to recognize Psi Chi faculty advisors for their outstanding service to the chapter and to Psi Chi.

Regional Chapter Awards | Dec 1
The Psi Chi Regional Chapter Awards provide annual recognition for up to two chapters in each region that best achieve Psi Chi’s purpose. Each winning chapter receives a check for $500 and a plaque to display in the winning chapter’s department. The awards are intended to perpetuate the chapters, to identify chapters as role models for others, and to promote the purposes of Psi Chi.

Regional Faculty Advisor Awards | Dec 1
This award is presented annually to one Psi Chi faculty advisor from each region who best achieves Psi Chi’s purpose. The award is intended to recognize and rewardPsi Chi advisors who are actively involved in chapter programs. The winning faculty advisor from each region will receive $500 and a plaque.

Cousins Award | Feb 1
The Psi Chi/Ruth Hubbard Cousins Chapter Award is presented annually to the chapter that best achieves Psi Chi’s purpose. The winning chapter receives: (1) a check for $3,500, (2) travel expenses for one chapter officer to attend the APA/Psi Chi Society Annual Convention to receive the award, and (3) a plaque to display in the winning chapter’s department.

Website Awards | Feb 1
These awards are presented annually to three chapters with websites that are innovative, aesthetically pleasing, and that advance or support Psi Chi’s purpose. Winning chapters will receive awards of $200 each.

Kay Wilson Leadership Award | April 1
The Psi Chi/Kay Wilson Leadership Award for Outstanding Chapter Presidents is presented annually to the one chapter president who demonstrates excellence in leadership of the local chapter. The winning Psi Chi chapter officer receives: (1) a $500 cash award, (2) travel expenses for one chapter president to attend and make a short presentation at the APA/Psi Chi Society Annual Convention to receive the award, and (3) an engraved plaque commemorating the award.

Model Chapter Awards | June 30
Model Chapter Awards of $1,000 each are presented annually to recognize and reward Psi Chi chapters that consistently maintain outstanding records of membership inductions, chapter correspondence, service projects, and other criteria associated with being an outstanding chapter. All chapters submitting evidence of meeting these criteria are designated as winners.